



66	71	2.8	188914	9	AC019118	Homo sapi
67	71	2.8	189695	2	AC130472	Homo sapi
68	71	2.8	197709	2	AC073641	Homo sapi
69	71	2.8	198388	2	AC073426	Homo sapi
70	71	2.8	198398	2	AC022219	Homo sapi
71	71	2.8	198748	2	AC016719	Homo sapi
72	71	2.8	200000	2	AP000494	Homo sapi
73	71	2.8	200885	2	AC012460	Homo sapi
74	71	2.8	209382	2	AC005071	Homo sapi
75	71	2.8	211945	2	AC006238	Homo sapi
76	71	2.8	215004	2	AC068146	Homo sapi
77	71	2.8	285886	2	AL391377	Homo sapi
78	71	2.8	303189	2	AC108728	Homo sapi
79	71	2.8	340000	2	AP001712	Homo sapi
80	70	2.7	104972	9	AC025176	Homo sapi
81	70	2.7	119185	9	AL157699	Homo sapi
82	70	2.7	135752	9	AC073465	Homo sapi
83	70	2.7	146802	9	AC025186	Homo sapi
84	70	2.7	177283	2	AC040942	Homo sapi
85	70	2.7	180948	2	AC099516	Homo sapi
86	70	2.7	183584	9	AC014399	Homo sapi
87	70	2.7	208561	9	AC012492	Homo sapi
88	70	2.7	211063	2	AC023206	Homo sapi
89	69	2.7	30257	2	AL603952	Homo sapi
90	69	2.7	42213	9	AC010527	Homo sapi
91	69	2.7	44735	9	AC092315	Homo sapi
92	69	2.7	48623	2	AC130469	Homo sapi
93	69	2.7	49372	9	AC109812	Homo sapi
94	69	2.7	60310	9	AL391647	Homo sapi
95	69	2.7	63753	2	AC090089	Homo sapi
96	69	2.7	89666	2	AC022403	Homo sapi
97	69	2.7	90727	9	AC015779	Homo sapi
98	69	2.7	93238	2	AC022713	Homo sapi
99	69	2.7	98261	9	HS1054C24	Human DNA
100	69	2.7	98261	9	HS1054C24	Human DNA

## ALIGNMENTS

RESULT 1  
LOCUS AX452880 2567 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 1 from Patent WO0242457.  
ACCESSION AX452880  
VERSION AX452880.1 GI:21712520  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Chang, H., Yang, W.P., Wu, Y., Whitney, G.S., Perez-Villar, J.J. and  
Kanner, S.B.  
TITLE Cloning and expression of human slap-2: a novel sh2/bh3  
domain-containing human slap homologue having immune cell-specific  
expression  
JOURNAL Patent: WO 0242457-A 1 30-MAY-2002;  
Bristol-Myers Squibb Co. (US)  
FEATURES  
source location/Qualifiers  
1. 2567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 611 a 741 c 666 g 549 t  
ORIGIN

Query Match 100.0%; Score 2567; DB 6; Length 2567;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCACGGCTCCGGTCGAGCTAGAGCTCCAGAGACCCACGCTGTGTCTGTGTGACAGA 60  
Db 1 CCCACGGCTCCGGTCGAGCTAGAGCTCCAGAGACCCACGCTGTGTCTGTGTGACAGA 60

Qy 61 GCTCAAGAGGCGCTGGGCTTCCCTCCGCTGCTGGCTGTGGAGGGTTCCCAAGT 120  
Db 61 GCTCAAGAGGCGCTGGGCTTCCCTCCGCTGCTGGCTGTGGAGGGTTCCCAAGT 120  
Qy 121 CCAGAAATCCCTTAAGAGAGCATGGGGAGCTATTCATCCCTGGTGTACAACTGTGACTG 180  
Db 121 CCAGAAATCCCTTAAGAGAGCATGGGGAGCTATTCATCCCTGGTGTACAACTGTGACTG 180  
Qy 121 CCAGAAATCCCTTAAGAGAGCATGGGGAGCTATTCATCCCTGGTGTACAACTGTGACTG 180  
Db 121 CCAGAAATCCCTTAAGAGAGCATGGGGAGCTATTCATCCCTGGTGTACAACTGTGACTG 180  
Qy 181 CAGACAGATGTGAGCTACCAACCAACCTAGCTCTCCCTGAAGACTCTCCAGAGC 240  
Db 181 CAGACAGATGTGAGCTACCAACCAACCTAGCTCTCCCTGAAGACTCTCCAGAGC 240  
Qy 241 TGAAGAGATCTGGGTGTCTTGAAGCCAAAGAGACATGGGACATTTCCAGAGGGCCCCCA 300  
Db 241 TGAAGAGATCTGGGTGTCTTGAAGCCAAAGAGACATGGGACATTTCCAGAGGGCCCCCA 300  
Qy 301 AAGCCTTAACCTGTCCAGCAGAGCATGCTCTCAGCAGAGCTGTCTTCCAGAGCTTTG 360  
Db 301 AAGCCTTAACCTGTCCAGCAGAGCATGCTCTCAGCAGAGCTGTCTTCCAGAGCTTTG 360  
Qy 361 ATGACAAACCAATTTCCCGATGATGTCTTGAAGCTCTCTGAGGAAATGGGA 420  
Db 361 ATGACAAACCAATTTCCCGATGATGTCTTGAAGCTCTCTGAGGAAATGGGA 420  
Qy 421 AGCTGCCAGCAGAAATCTCTGCGCAAGCCCAAGCTTGAATTCCTCTGCAAGGC 480  
Db 421 AGCTGCCAGCAGAAATCTCTGCGCAAGCCCAAGCTTGAATTCCTCTGCAAGGC 480  
Qy 481 CAGGAGCTGTGACATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Db 481 CAGGAGCTGTGACATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Qy 541 TTCCCGCAGTGGCCCGCGGAGCTGTGCTGAGACTCGGGAGGCATTGACATCTGTC 600  
Db 541 TTCCCGCAGTGGCCCGCGGAGCTGTGCTGAGACTCGGGAGGCATTGACATCTGTC 600  
Qy 601 TCTGAGATGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTAATC 660  
Db 601 TCTGAGATGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTAATC 660  
Qy 661 CCCACGGCTCCAGTGGCCAAATGCTTCCCATGGGTGGCTGTATGAGGGCTTGAGCAGGAG 720  
Db 661 CCCACGGCTCCAGTGGCCAAATGCTTCCCATGGGTGGCTGTATGAGGGCTTGAGCAGGAG 720  
Qy 721 AAACAGAGAGAGAGCTGTGTTTACCTGGAGAACCTGGAGGGGCTTCCCTCATCCGGAG 780  
Db 721 AAACAGAGAGAGAGCTGTGTTTACCTGGAGAACCTGGAGGGGCTTCCCTCATCCGGAG 780  
Qy 781 AGCAGACAGAGAGAGGCTTTTACTCTGTGACGTCCGCTCAGCCGCTGCAATCTCTG 840  
Db 781 AGCAGACAGAGAGAGGCTTTTACTCTGTGACGTCCGCTCAGCCGCTGCAATCTCTG 840  
Qy 841 GACCGATCAGACATACAGAGATCCATGCTTGAACATGAGAGAGAGAGAGAGAGAGAG 900  
Db 841 GACCGATCAGACATACAGAGATCCATGCTTGAACATGAGAGAGAGAGAGAGAGAGAG 900  
Qy 901 CGCTCACTTCCCTTCACTCAGAGCCCTGTGTGACCAATTAATCTGAGTGGCGAGTAC 960  
Db 901 CGCTCACTTCCCTTCACTCAGAGCCCTGTGTGACCAATTAATCTGAGTGGCGAGTAC 960  
Qy 961 ATGCTGCTACTCAAGAGAGCTGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 ATGCTGCTACTCAAGAGAGCTGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Qy 1021 GATATACCTTAACCTGTGACTGTGAGAGAGACCACTCAATGAGAAAGAGTGAACAGC 1080  
Db 1021 GATATACCTTAACCTGTGACTGTGAGAGAGACCACTCAATGAGAAAGAGTGAACAGC 1080  
Qy 1081 TCCCTCTGTTTCTGAGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 1081 TCCCTCTGTTTCTGAGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140



BASE COUNT	ORIGIN
599 a	662 c 618 g 536 t

Query Match	72.5%	Score 1860;	DB 9;	Length 2415;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 2130; Conservative	0;	Mismatches 3;	Indels 1;	Gaps 1;

QY 415 ATGGAAAGCTCCCGACGAAAGAAATCTGCGCAAGCCCAAGCTTGAAGTCTCTGTC 474  
 DB 61 ATGGAAAGCTCCCGACGAAAGAAATCTGCGCAAGCCCAAGCTTGAAGTCTCTGTC 120  
 QY 475 CAAGGCCAGGAACCTGTGACCTAGAGACAGAGAGACAAAGCCACAGCCGTGACCTG 514  
 DB 121 CAAGGCCAGGAACCTGTGACCTAGAGAGAGAGAGACAAAGCCACAGCCGTGACCTG 180  
 QY 535 GGCAGTTTCCCGGACGAGTGGCCGGCCAGAGCTGTGGCTGAGAGCTCGGGAGGCAATTAGC 594  
 DB 181 GGCAGTTTCCCGGACGAGTGGCCGGCCAGAGCTGTGGCTGAGAGCTCGGGAGGCAATTAGC 240  
 QY 595 ATGCTCTGAGGATGAGACTGTGTGACGGTGTCTGTGAAGTCTCAGGACAGAGATGAT 654  
 DB 241 ATGCTCTGAGGATGAGACTGTGTGAGGGGTGCTGTGAAGTCTCAGGACAGAGATGAT 300  
 QY 655 AACATCCCAAGGCTCAGTGGCCCAAAAGTCTCCATGGAGTGGCTGTATAGAGGACCTGAGC 714  
 DB 301 AACATCCCAAGGCTCAGTGGCCCAAAAGTCTCCATGGAGTGGCTGTATAGAGGACCTGAGC 360  
 QY 715 AGGAGAAAGCAGAGGAATCTGTGTGTTACTCTGGAAACCTTGAGAGGACCTTCTCATC 774  
 DB 361 AGGAGAAAGCAGAGGAATCTGTGTGTTACTCTGGAAACCTTGAGAGGACCTTCTCATC 420  
 QY 775 CGGAGAGCCAAACAGAGAGAGGCTCTTACTCTGTGATCGGCTCAGCGCCGCTTGCA 834  
 DB 421 CGGAGAGCCAAACAGAGAGAGGCTCTTACTCTGTGATCGGCTCAGCGCCGCTTGCA 480  
 QY 835 TCTGTGGAGCCGATCAGACACTACAGAGATCCATGCTTGACAAATGCTGTGCTTGATC 894  
 DB 481 TCTGTGGAGCCGATCAGACACTACAGAGATCCATGCTTGACAAATGCTGTGCTTGATC 540  
 QY 895 TCACCGGCGCTCACTTCCCTCACTCCAGGCGCTGTGTGACAAATTACTGTAGCTGAGC 954  
 DB 541 TCACCGGCGCTCACTTCCCTCACTCCAGGCGCTGTGTGACAAATTACTGTAGCTGAGC 600  
 QY 955 GATGACATCTGCTGCTACTCAGAGAGCCCTGTGCTGTGACAGAGGCTGACCCGCTCCT 1014  
 DB 601 GATGACATCTGCTGCTACTCAGAGAGCCCTGTGCTGTGACAGAGGCTGACCCGCTCCT 660  
 QY 1015 GGCAGAGATATACCCCTACTGTGACTGTGACAGAGAGACCAACTCAACTGAGAAAGACTG 1074  
 DB 661 GGCAGAGATATACCCCTACTGTGACTGTGACAGAGAGACCAACTCAACTGAGAAAGACTG 720  
 QY 1075 GACAGCTCCCTCTCTGTTTCTGAAAGCTGCCACAGGGAGAGTCTCTTCTCAGTGAAGGT 1134  
 DB 721 GACAGCTCCCTCTCTGTTTCTGAAAGCTGCCACAGGGAGAGTCTCTTCTCAGTGAAGGT 780  
 QY 1135 CTCGGAGAGTCCCTCAGCTTATCATCAGCTGAAATGACAGAGGTGCTCTTTGGATGAT 1194  
 DB 781 CTCGGAGAGTCCCTCAGCTTATCATCAGCTGAAATGACAGAGGTGCTCTTTGGATGAT 840  
 QY 1195 GCTTAGGCTCAAAAGAGAGGCCAAAGAGAAACAAAGGCTGCACACTGAAACCCCAATT 1254  
 DB 841 GCTTAGGCTCAAAAGAGAGGCCAAAGAGAAACAAAGGCTGCACACTGAAACCCCAATT 900  
 QY 1255 CAGGCTCTGTGGGACCCCAAGAGGCAAGGCTGTGACTCAGGAGAGGAGGATGGGACAG 1314  
 DB 901 CAGGCTCTGTGGGACCCCAAGAGGCAAGGCTGTGACTCAGGAGAGGAGGATGGGACAG 960  
 QY 1315 AGGTGACTTAGGGTCCACTGTGACCTTGTCTTTCTTCTTGAAGCTTGAAGTCA 1374

Db	961	AGTGCATCTTAAGGATCCACCTGACCTTGCTTTTCCCTCTTAGGCCCTTAGAAGTCA	1023
Qy	1375	CCTACTTCTTCCAGTGGCAATATCCCACTGCGACCTCTAATGGAGTGCAGAAAGGT	1434
Db	1021	CCCAATCTCTTCCAGTGCATGATATCCCACTGCACTCTAGTGGAGTGCAGAAAGGT	1080
Qy	1435	GGGACCAAGGCGCAGGATTCAAAAAAGAAATTAAGCTCTTGGGGGGTTGTACTAGTTAG	1494
Db	1081	GGGACCAAGGCGCAGGATTCAAAAAAGAAATTAAGCTCTTGGGGGGTTGTACTAGTTAG	1140
Qy	1495	TTCTTGAGTTGGGGTTCCAGTACCATGCGATGAGGCGCTGCTGTTAGGCCATTTAC	1554
Db	1141	TTCTTGAGTTGGGGTTTCAGTACCATGCGATGAGGCGCTGCTGTTAGGCCATTTAC	1200
Qy	1555	ATCCCAACATTAACAGAGCCCCACCAAGATGAAACAACCCCTAGAGTCAACAGA	1614
Db	1201	ATCCCAACATTAACAGAGCCCCACCAAGATGAAACAACCCCTAGAGTCAACAGA	1260
Qy	1615	AAGTCATTTTCAAAAAATCTCAAGTCTGTTGAGACACACACATATCTCAAGAGTAG	1674
Db	1261	AAGTCATTTTCAAAAAATCTCAAGTCTGTTGAGACACACACATATCTCAAGAGTAG	1320
Qy	1675	GACTGTGCTTAGAAGGAAAGAAAGTGAATGATGTCTTACCTTAGACAGATCTT	1734
Db	1321	GACTGTGCTTAGAAGGAAAGAAAGTGAATGATGTCTTACCTTAGACAGATCTT	1380
Qy	1735	GGATGGTCAAGGCTCTAATGTACCTCGACAGAAAGAAAGCTTCCGACACTAGGT	1794
Db	1381	GGATGGTCAAGGCTCTAATGTACCTCGACAGAAAGAAAGCTTCCGACACTAGGT	1440
Qy	1795	CCTCAATGTCCCCATTGAGAACAAAGCCCAAGCTCTTTTCTTTTTTTGAAACGG	1854
Db	1441	CCTCAATGTCCCCATTGAGAACAAAGCCCAAGCTCTTTTCTTTTTTTGAAACGG	1499
Qy	1855	AGCTCTGCGCTGTGGCCATCTGAGATGAGCATGAGCATCTCAAGCTCACTCAACTC	1914
Db	1500	AGCTCTGCGCTGTGGCCATCTGAGATGAGCATGAGCATCTCAAGCTCACTCAACTC	1559
Qy	1915	CATCTCTGGAATTCAAACAAATCTCTGCTCCAGCTCCCAAGTAAGCTGGATTAAGC	1974
Db	1560	CATCTCTGGAATTCAAACAAATCTCTGCTCCAGCTCCCAAGTAAGCTGGATTAAGC	1619
Qy	1975	GTACACACACATGCTCGGCTAATTTTTTTGTAATTTTGTAGACATGAGGTTTACACA	2034
Db	1620	GTACACACACATGCTCGGCTAATTTTTTTGTAATTTTGTAGACATGAGGTTTACACA	1679
Qy	2035	TTGGCCAGGCTGTGTCAATCTCTGACCTCAAGTATCCACCACTTGSCCTCCGCAAA	2094
Db	1680	TTGGCCAGGCTGTGTCAATCTCTGACCTCAAGTATCCACCACTTGSCCTCCGCAAA	1739
Qy	2095	GTGCTGGGATTAGAGGTGTGAGCCACGGACCCACAGCTAGCTCTCAGATCTTAATTCAT	2154
Db	1740	GTGCTGGGATTAGAGGTGTGAGCCACGGACCCACAGCTAGCTCTCAGATCTTAATTCAT	1799
Qy	2155	TTTGTGGTTTACAATTCCTGTAGCACTAGGCGTTGCCATCTTGTCGCGAATAAAAAATA	2214
Db	1800	TTTGTGGTTTACAATTCCTGTAGCACTAGGCGTTGCCATCTTGTCGCGAATAAAAAATA	1859
Qy	2215	ACACCTCTTAAGTCTAGCACTGCAATGAGGCGCAGGACCTTAGTGTGGCAGGGGCA	2274
Db	1860	ACACCTCTTAAGTCTAGCACTGCAATGAGGCGCAGGACCTTAGTGTGGCAGGGGCA	1919
Qy	2275	TCAGAAAGTGTAAAGCCCTCTCTCAATATGCAAGAGGACACACAGCTTACACAAA	2334
Db	1920	TCAGAAAGTGTAAAGCCCTCTCTCAATATGCAAGAGGACACACAGCTTACACAAA	1979
Qy	2335	TCCAGCCCTTGATTTTCCCTGTGCTTCAATAACAAAGAGGTCTGTGATCTGCTTAA	2394
Db	1980	TCCAGCCCTTGATTTTCCCTGTGCTTCAATAACAAAGAGGTCTGTGATCTGCTTAA	2039
Qy	2395	GGATCTAAGGAGAGAAAGAAAGAGGATGGGGTGGAGAGCAACCCCTCAATGCTCTTAC	2454
Db	2040	GGATCTAAGGAGAGAAAGAAAGAGGATGGGGTGGAGAGCAACCCCTCAATGCTCTTAC	2099



Qy	Db	Qy	Db
2455	TGGTCCCAAGCTACAGGTGGGGTGGGAAAGCTTTATCAGATATCATCAACAGTTC	2514	TGGTCCCAAGCTACAGGTGGGGTGGGAAAGCTTTATCAGATATCATCAACAGTTC
2100	TGGTCCCAAGCTACAGGTGGGGTGGGAAAGCTTTATCAGATATCATCAACAGTTC	2155	TGGTCCCAAGCTACAGGTGGGGTGGGAAAGCTTTATCAGATATCATCAACAGTTC
2515	AATTAAAGATTGATTTATTCAGATATGCGAAA	2548	AATTAAAGATTGATTTATTCAGATATGCGAAA
2160	AATTAAAGATTGATTTATTCAGATATGCGAAA	2193	AATTAAAGATTGATTTATTCAGATATGCGAAA

RESULT 3	
HSDJ977B1/c	
LOCUS	HSDJ977B1 145068 bp DNA linear PRI 20-JUL-2001
DEFINITION	Human DNA sequence from clone R5-977B1 on chromosome 20 Contains

**KEYWORDS** HTG; CpG island; heterogeneous ribonucleoprotein; KIAA0964; myosin regulatory light chain; MRL2; RAB5-interacting protein; SH2 domain; TGIF2; transcription factor; tyrosine kinase.

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1 (bases 1 to 145068)
<i>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.</i>	

**TITLE** Direct Submission  
**SUBMITTED** 19-JUL-2001 **SENDER** Centre, Hinxton, Cambridgeshire,  
**JOURNAL** CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
**REQUESTER** clonerequest@sanger.ac.uk  
**COMMENT** On Jul 28, 2000 this sequence replaced gi:5924017.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that that variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary access numbers given in the feature table with their source databases: Emi; EMBL; SW; SWISSPROT; Tr; TRMBL; Mp; WOMMPP; Information on the WOMMPP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wommpp](http://www.sanger.ac.uk/Projects/C_elegans/wommpp) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

This sequence is the entire insert of clone RP5-977B1. The true 1 kb end of clone RP3-460J8 is at 63345 in this sequence. The true 1 kb end of clone CMD-218279 is at 62218 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry covered by high quality data (i.e., phred quality  $\geq 30$ ); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-977B1 is from the library RPc1-5 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2.

```

repeat_region /clone="PP5-977B1"
                    /clone_lib="RPcrl-5"
568..875
/note="AlusX repeat: matches 1..311 of consensus"
misc_feature 1024..1580
/note="match: GSS: Em:AQ620498"
misc_feature 1047..1629
/note="match: GSS: Em:AQ779162"
repeat_region 1631..1923
/note="AlusG repeat: matches 1..294 of consensus"
repeat_region 2776..2842
/note="L2 repeat: matches 2257..2423 of consensus"
repeat_region 3891..3949
/note="MIR repeat: matches 34..90 of consensus"
repeat_region 4458..4611
/note="MIR repeat: matches 67..231 of consensus"
repeat_region 5133..5549
/note="LM2D repeat: matches 5924..6340 of consensus"
repeat_region 5550..5660
/note="PRAM C repeat: matches 1..132 of consensus"
repeat_region 5703..5890
/note="AlusG repeat: matches 120..308 of consensus"
repeat_region 5923..6012
/note="LMC/D repeat: matches 5432..5528 of consensus"
repeat_region 6588..6889
/note="AlusP repeat: matches 1..303 of consensus"
gene join(7423..7784,9960..10039,10306..10394,10917..11329,
34968..35059,36599..36724,37531..39351)
(gene="KIAA0964"
join(c7423..7784,9960..10039,10917..11329,34968..35059,
36599..36724,37531..39351))
/gene="KIAA0964"
/product="cdg977B1.6.2 (KIAO964 [ortholog of rat
PSD-95/SAP90-associated protein 4])"
/note="match: CDNA5: Em:AB002277 Em:AB000276
Em:AB000275 Em:AB023181 Em:AF009204 Em:AB003594 Em:U67988
Em:U67137 Em:U67138 Em:U67987 Em:U67139 Em:AT131778
match: BSRs: Em:A1811867 Em:A1820916 Em:AA496750
Em:AA647552 Em:A1747584 Em:AA931116 Em:A1887761
Em:A154844 Em:A1918247 Em:H05453 Em:A1202706 Em:A197978
Em:A1914247 Em:AA611645 Em:AA760107 Em:AA591960
Em:AA469640 Em:AA190916 Em:T07912 Em:AA552402 Em:A1748854
Em:A1316293 Em:A1840413 Em:A17317 Em:AA570114
Em:AA870776 Em:AA144662 Em:A161315 Em:AA496605 Em:W62099
Em:AA43364 Em:AA874407 Em:A1986110 Em:AA496009
Em:AA953725 Em:AB95674 Em:A1493631 Em:AA760121
Em:AA050197 Em:R84650 Em:J35758 Em:F06905 Em:AA33898
Em:BS4451 Em:A1918247 Em:H05453 Em:W42663 Em:TV108
Em:AA824332 Em:R13930 Em:A194869 Em:AA861813 Em:AA602005
Em:A1201639 Em:W45180 Em:Z45916 Em:F12931 Em:A1936733
Em:F12936 Em:N36776 Em:H12916 Em:I13355 Em:A1983945
Em:A1126063 Em:T57939 Em:A1916248 Em:TV5144 Em:W07614
Em:A1972603 Em:AA434575 Em:A1500286 Em:A1501049
Em:A1334454 Em:A1128723 Em:A1719558 Em:AO301035 Em:T08824
Em:AS2662 Em:RS1956 Em:R6064 Em:Z42446 Em:TS9317
Em:R96101 Em:H41361 Em:AA064703 Em:AA25888 Em:AL049119
Em:AA222525 Em:D28675 Em:A1652021 Em:AA0882413 Em:H05501
Em:AA4297 Em:AA046329 Em:N45937 Em:AA577212 Em:AA610855
Em:A1017860 Em:W58510 Em:T08842 Em:A1057148 Em:N2846"
/evidence=not experimental
join(c7423..7784,9960..10039,10917..11329,34968..35059,
36599..36724,37531..37749)
/gene="KIAA0964"
/note="isoform 2
continues in cd996C2 (AL445705) and bk2182L9 (AL390374)
match: proteins: Tr:O9J2H0"

```



Query Match	44.1%	Score 1132;	DB 6;	Length 1183;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1182;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;
QY	18	ACGTAAAGCTCCAAAGAACCCGAGGCGCTGTGTCTCTGTGACAGAGCTCAAAAGGCCCTGGG	77	
Db	1	AGCTAAAGCTCCAAAGAACCCGAGGCGCTGTGTCTGTGACAGAGCTCAAAAGGCCCTGGG	60	
QY	78	CGTTCCCTCCCTGGGCTGGGCTGTGGCTTTGGGAGGGGTTCCCGAGTCCAGAACTCCCTAAGAG	137	
Db	61	CGTTCCCTCCCTGGGCTGGGCTGTGGCTTTGGGAGGGGTTCCCGAGTCCAGAACTCCCTAAGAG	120	
QY	138	CATGGGCGAGCTGATTCATCCCTGTGTGTCAAACTGTGACTGACGTGAACAGATGCTGAGCT	197	
Db	121	CATGGGCGAGCTGATTCATCCCTGTGTGTCAAACTGTGACTGACGTGAACAGATGCTGAGCT	180	
QY	198	ACCCAAACCAACACTAGCTCTCTCCCTGAAGATCTCTCCAGAGGCTGAGAGATTTGGGGTG	257	
Db	181	ACCCAAACCAACACTAGCTCTCTCCCTGAAGATCTCTCCAGAGGCTGAGAGATTTGGGGTG	240	
QY	258	TCCTAGGACCAAGACATGTGCAAGATTTCCAAAGAGGCCCCCAAGCCTTAACCTGTTC	317	

Db	241	TCCTAAGGACCAAGGACACTGGACAGACTTCCAGAAAGGAGCCCCCAAGGACCTTAACCTGTGTCA	300
Qy	318	GCACAGACATGGGCTCTCAGAGAGAGCTGCTTCCCAAGCCTTTGATGTACAAACAAATTTCC	377
Db	301	GCAGAGACATGGGTCTCAGAGAGAGCTGCTTCCCAAGCCTTTGATGTACAAACAAATTTCC	360
Qy	378	CTCGATGATGTGCTTCTGAGTGTCTCTGTAGAGAACAAATGGAAAGTGTGCTCCAGCAGAG	437
Db	361	CTCGATGATGTGCTTCTGAGTGTCTCTGTAGAGAACAAATGGAAAGTGTGCTCCAGCAGAG	420
Qy	438	AAATCTCTGCCAACCCCAAGCTGATGTTCCCTGTGCAAGAGCCAGGAGACCTGTGACAT	497
Db	421	AAATCTCTGCCAACCCCAAGCTGATGTTCCCTGTGCAAGAGCCAGGAGACCTGTGACAT	480
Qy	498	GGAGCAGAGAGAACAGAGCCACAGCCGTGGCCTTGAGCAAGTTTCCGAGAGTGGCC	557
Db	481	GGAGCAGAGAGAACAGAGCCACAGCCGTGGCCTTGAGCAAGTTTCCGAGAGTGGCC	540
Qy	558	GGCCGACCTGTGTGAGAGACTGGGGAGAACCTTGACATGCTCTGTAGAGATGAGAGCTG	617
Db	541	GGCCGACCTGTGTGAGAGACTGGGGAGAACCTTGACATGCTCTGTAGAGATGAGAGCTG	600
Qy	618	GTGACAGGTGCTGTCTGAGTCTCAGGAGAGATTAACATCCCAAGCTCAGGTGGC	677
Db	601	GTGACAGGTGCTGTCTGAGTCTCAGGAGAGATTAACATCCCAAGCTCAGGTGGG	660
Qy	678	CAAGGTCTCCATGGGTGGCTGTATATAGAGGCTGTAGACAGGAGAAACAGAGAACTGCT	737
Db	661	CAAGGTCTCCATGGGTGGCTGTATATAGAGGCTGTAGACAGGAGAAACAGAGAACTGCT	720
Qy	738	GTGTGTACCTTGAGAAACCTGTGAGAGGACCTTCTTCATCCGAGAGACCAAGCAGAGAGG	797
Db	721	GTGTGTACCTTGAGAAACCTGTGAGAGGACCTTCTTCATCCGAGAGACCAAGCAGAGAGG	780
Qy	798	CTCTTACTCTGTGACGTCCGCTCAGCCGACCTGTGACCTGTGGAGCCGATCAGACACTA	857
Db	781	CTCTTACTCTGTGACGTCCGCTCAGCCGACCTGTGACCTGTGGAGCCGATCAGACACTA	840
Qy	858	CAGATCTCACTGCTTTGACAAATGATGATCTGATCATCCGCGCTCACTTCCCTCTC	917
Db	841	CAGATCTCACTGCTTTGACAAATGATGATCTGATCATCCGCGCTCACTTCCCTCTC	900
Qy	918	ACTCCAGAGCCCTGTGTGAGACCTTAATCTGTGATGGGAGATGACATCTGCTGCTACTCA	977
Db	901	ACTCCAGAGCCCTGTGTGAGACCTTAATCTGTGATGGGAGATGACATCTGCTGCTACTCA	960
Qy	978	GGAGCCCTGTGTCTTGACAGAGGCTGGCCGCTTCCGTGGCAAGATTAACCTTAAGCTGT	1037
Db	961	GGAGCCCTGTGTCTTGACAGAGGCTGGCCGCTTCCGTGGCAAGATTAACCTTAAGCTGT	1020
Qy	1038	GACTGTGCTCAGAGACACCACTCAACTGTGAAAGAGCTGACAGCTCCCTCTGTTTCTGA	1097
Db	1021	GACTGTGCTCAGAGACACCACTCAACTGTGAAAGAGCTGACAGCTCCCTCTGTTTCTGA	1080
Qy	1098	ACCTGTCCACAGGGAGAGAGTCTTTTCTCAGTGAAGGTCTTCGGAGAGTCCCTCAAGTTCTA	1157
Db	1081	ACCTGTCCACAGGGAGAGAGTCTTTTCTCAGTGAAGGTCTTCGGAGAGTCCCTCAAGTTCTA	1140
Qy	1158	CATCAGCCTGAATGACAGAGGTCTCTCTTTGGATGATGCTTAG	1200
Db	1141	CATCAGCCTGAATGACAGAGGTCTCTCTTTGGATGATGCTTAG	1183
RESULT 5			
AX443135/c			
LOCUS	AX443135	1183 bp	DNA linear PAT 02-JUL-2002
DEFINITION	Sequence 76 from Patent W00216599.		
ACCESSION	AX443135		
VERSION	AX443135.1	GI:21690556	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

REFERENCE  
1 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
AUTHORS  
Burgess, C.E., Conley, P.B., Grosse, W.M., Hart, M., Kekula, R.,  
Shinkov, R.A., Speck, K.A., Szekeres, E.S., Tomlinson, J.E.,  
Topper, J.N. and Yang, R.B.  
TITLE  
Protein and nucleic acid encoding same  
JOURNAL  
Patent: WO 0216599-A 76 28-FEB-2002.  
Curagen Corporation (US); COR THERAPEUTICS, INC. (US)  
FEATURES  
source  
1. 1183  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 240 a 333 c 359 g 251 t  
ORIGIN

Query Match 44.1%; Score 1132; DB 6; Length 1183;  
Beet Local Similarity 99.9%; Pred. No. 0;  
Matches 1182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 18 AGTAGAGCTCAAGAGACCCAGCCTGTCTCTGTGACAGAGCTCAAGAGGCGCTGG 77  
Db 1183 AGTAGAGCTCAAGAGACCCAGCCTGTCTCTGTGACAGAGCTCAAGAGGCGCTGG 1124

QY 78 CCTTCCCTCCCTGGCTGGCTGTCTGTGGAGGCTTCCCACTCAGATCCCTAAGAG 137  
Db 1123 CCTTCCCTCCCTGGCTGGCTGTCTGTGGAGGCTTCCCACTCAGATCCCTAAGAG 1064

QY 138 CATGGGAGCTGATCATCCCTGTGTACAACTGCTGACAGAGATGCTGAGCT 197  
Db 1063 CATGGGAGCTGATCATCCCTGTGTGTACAACTGCTGACAGAGATGCTGAGCT 1004

QY 138 ACCCAACCAACCTTACCTCTCCCTGAAAGATCTCCAGAGCTGAGAGATTTGGGTG 257  
Db 1003 ACCCAACCAACCTTACCTCTCCCTGAAAGATCTCCAGAGCTGAGAGATTTGGGTG 944

QY 258 TCTTAGAGCCAGAGACACTGAGACTTCCAGAGAGGCGCCCAAGGCTTAACTGTGCA 317  
Db 943 TCTTAGAGCCAGAGACACTGAGACTTCCAGAGAGGCGCCCAAGGCTTAACTGTGCA 884

QY 318 GCCAGAGACCTGCTGACAGAGAGCTCTTCCCAAGGCTTGTATGACAACTTTCC 377  
Db 883 GCCAGAGACCTGCTGACAGAGAGCTCTTCCCAAGGCTTGTATGACAACTTTCC 824

QY 378 CTGAGTATGTCTGTAGTGTCTCTGTGAGGACAAATGGAGATGTGCTCCAGAGAG 437  
Db 823 CTGAGTATGTCTGTAGTGTCTCTGTGAGGACAAATGGAGATGTGCTCCAGAGAG 764

QY 438 AAAATCTCTGCCAAGCCAGCTTGAATCTCTGTCCAGAGGAGAGAGCTGATCAT 497  
Db 763 AAAATCTCTGCCAAGCCAGCTTGAATCTCTGTCCAGAGGAGAGAGCTGATCAT 704

QY 498 GGAAGCAGAGAGAGAGAGAGGACAGAGGCTGAGGAGCTTCCGAGAGAGTGGCC 557  
Db 703 GGAAGCAGAGAGAGAGAGAGGACAGAGGCTGAGGAGCTTCCGAGAGAGTGGCC 644

QY 558 GGGCGAGCTGTGCTGAGAGCTGGGAGAGCAATGCTCTCTGAGAGATGAGAGT 617  
Db 643 GGGCGAGCTGTGCTGAGAGCTGGGAGAGCAATGCTCTCTGAGAGATGAGAGT 584

QY 618 GTGAGAGCTGTCTGAGAGCTGAGAGAGATTAATCCAGAGCTCAGCTGGC 677  
Db 583 GTGAGAGCTGTCTGAGAGCTGAGAGAGATTAATCCAGAGCTCAGCTGGC 524

QY 678 CAAGTCTCCATGAGGTGTGTGAGAGGCTGAGAGAGAGAGAGAGAGAGAGTCT 737  
Db 523 CAAGTCTCCATGAGGTGTGTGAGAGGCTGAGAGAGAGAGAGAGAGAGTCT 464

QY 738 GTTGTATCTGGAG 797  
Db 463 GTTGTATCTGGAG 404

QY 798 CTCTTACTCTGTGAGTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 857

Db 403 CTCTTACTCTGTGAGTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 344

QY 858 CAGAGATCACTGCTGAG 917

Db 343 CAGAGATCACTGCTGAG 284

QY 918 ACTTCAG 977

Db 283 ACTTCAG 224

QY 978 GAGAGCTGTCTGTGAG 1037

Db 223 GAGAGCTGTCTGTGAG 164

QY 1038 GACTGTGAG 1097

Db 103 GACTGTGAG 44

QY 1098 AGCTGCAAG 1157

Db 103 AGCTGCAAG 44

QY 1158 CATCAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

Db 43 CATCAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1

RESULT 6  
AC026539 145833 bp DNA linear HTG 27-APR-2000  
LOCUS  
DEFINITION  
Homo sapiens chromosome 20 clone RP11-712N14 map 20, WORKING DRAFT  
SEQUENCE, 37 unordered pieces.  
AC026539  
AC026539 2 GI:7656813  
VERSION  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 145833)  
AUTHORS  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boculavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choe, Y., Colangelo, M., Collins, S.,  
Collins, A., Cooke, P., Doyle, M., Ferrel, P., Fitzhugh, W., Gage, D.,  
Dodge, S., Domino, M., Doyle, M., Ferrel, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hago, B., Heald, J., Horton, L.,  
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marcini, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,  
Melton, D., Menus, L., Mihova, T., Miranda, C., Mogen, J., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessier, S., Theodore, J., Tjelle, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,  
Young, G., Zaitoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 27, 2000 this sequence version replaced gt:7283243.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBC
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L7115
Center clone name: 712.N.14
----- Summary Statistics -----
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 125577 bases at least Q40
Consensus quality: 135703 bases at least Q30
Consensus quality: 139593 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 142233; sum-of-contigs
Quality coverage: 2.6 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1214: contig of 1214 bp in length
* 1215 1314: gap of 100 bp
* 1315 2673: contig of 1359 bp in length
* 2674 2773: gap of 100 bp
* 2774 4520: contig of 1747 bp in length
* 4521 4620: gap of 100 bp
* 4621 5961: contig of 1341 bp in length
* 5962 6061: gap of 100 bp
* 6062 7719: contig of 1638 bp in length
* 7720 7819: gap of 100 bp
* 7820 9799: contig of 1980 bp in length
* 9800 9899: gap of 100 bp
* 9900 11434: contig of 1535 bp in length
* 11435 11534: gap of 100 bp
* 11535 14382: contig of 2848 bp in length
* 14383 14482: gap of 100 bp
* 14483 17148: contig of 2666 bp in length
* 17149 17248: gap of 100 bp
* 17249 19239: contig of 1991 bp in length
* 19240 19339: gap of 100 bp
* 19340 21102: contig of 1763 bp in length
* 21103 21202: gap of 100 bp
* 21203 23371: contig of 2169 bp in length
* 23372 23471: gap of 100 bp
* 23472 25782: contig of 2311 bp in length
* 25783 25882: gap of 100 bp
* 25883 28824: contig of 2942 bp in length
* 28825 28924: gap of 100 bp
* 28925 31619: contig of 2695 bp in length
* 31620 31719: gap of 100 bp
* 31720 34679: contig of 2960 bp in length
* 34680 34779: gap of 100 bp
* 34780 37527: contig of 2748 bp in length
* 37528 37627: gap of 100 bp
* 37628 40146: contig of 2519 bp in length
* 40147 40246: gap of 100 bp
* 40247 43743: contig of 3497 bp in length
* 43744 43843: gap of 100 bp
* 43844 47133: contig of 3290 bp in length
* 47134 47233: gap of 100 bp
* 47234 51023: contig of 3790 bp in length
* 51024 51123: gap of 100 bp
* 51124 54935: contig of 3812 bp in length
* 54936 55035: gap of 100 bp
* 55036 59553: contig of 4518 bp in length
* 59554 59653: gap of 100 bp

```

---

```

FEATURES
source
* 59654 64428: contig of 4775 bp in length
* 64429 64528: gap of 100 bp
* 64529 69211: contig of 4683 bp in length
* 69212 69311: gap of 100 bp
* 69312 72901: contig of 3590 bp in length
* 72902 73001: gap of 100 bp
* 73002 76724: contig of 3723 bp in length
* 76725 76824: gap of 100 bp
* 76825 81179: contig of 4355 bp in length
* 81180 81279: gap of 100 bp
* 81280 87009: contig of 5730 bp in length
* 87010 87109: gap of 100 bp
* 87110 90855: contig of 3746 bp in length
* 90856 90955: gap of 100 bp
* 90956 96520: contig of 5565 bp in length
* 96521 96620: gap of 100 bp
* 96621 102321: contig of 5701 bp in length
* 102322 102421: gap of 100 bp
* 102422 108293: contig of 5872 bp in length
* 108294 108393: gap of 100 bp
* 108394 116689: contig of 8286 bp in length
* 116690 116789: gap of 100 bp
* 116790 125264: contig of 8475 bp in length
* 125265 125364: gap of 100 bp
* 125365 136354: contig of 10990 bp in length
* 136355 136454: gap of 100 bp
* 136455 145833: contig of 9379 bp in length.
Location/Qualifiers
1. 145833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20"
/clone="RP11-712N14"
/clone_1fb="RP11-11 Human Male BAC"
1. 1214
/note="assembly_fragment"
1315. 2673
/note="assembly_fragment"
2774. 4520
/note="assembly_fragment"
4621. 5961
/note="assembly_fragment"
6062. 7719
/note="assembly_fragment"
7820. 9799
/note="assembly_fragment"
9900. 11434
/note="assembly_fragment"
11535. 14382
/note="assembly_fragment"
14483. 17148
/note="assembly_fragment"
17249. 19239
/note="assembly_fragment"
21103. 21202
/note="assembly_fragment"
21203. 23371
/note="assembly_fragment"
23472. 25782
/note="assembly_fragment"
25883. 28824
/note="assembly_fragment"
28925. 31619
/note="assembly_fragment"
31720. 34679
/note="assembly_fragment"
34780. 37527
/note="assembly_fragment"
37628. 40146
/note="assembly_fragment"
40247. 43743
/note="assembly_fragment"

```

QY	1076	ACAGCTCCCTCCGTGTTTCTGAACTCCGCAACGGGAGAGATCTCTTCTCAGTAGGGCTC	1135
Db	60571	AACACTCCCTCCGTGTTTCTGAACTCCGCAACGGGAGAGATCTCTTCTCAGTAGGGCTC	60512
QY	1136	TCCGGAGTCCCTCAGCTTCTACATCAGCCTGATTAAGAGAGCTGTCTCTTGGATGATG	1195
Db	60511	TCCGGAGTCCCTCAGCTTCTACATCAGCCTGATTAAGAGAGCTGTCTCTTGGATGATG	60452
QY	1196	CTTAGGCCCAAGAGAGAGCCAAAGGAAACCAAGGTGACACCTAGAACCCCAATTC	1255
Db	60451	CTTAGGCCCAAGAGAGAGCCCAAGGAAACCAAGGTGACACCTAGAACCCCAATTC	60392
QY	1286	AGCCTCTGGGGACCCCAAGAGGCAAGGCTGTGCACTCAGGGAGGGAGGGTGGACACAGA	1315
Db	60391	AGCCTCTGGGGACCCCAAGAGGCAAGGCTGTGCACTCAGGGAGGGAGGGTGGACACAGA	60332
QY	1316	GGTGCATCTTAGGGTCCCACTGTACCTCTGTCTTCTCTCTTAGACCTTGAAGTCAAC	1375
Db	60331	GGTGCATCTTAGGGTCCCACTGTACCTCTGTCTTCTCTCTTAGACCTTGAAGTCAAC	60272
QY	1376	CTACTTCTCTTCCAGTGCATATGCCACCTGGGACCTTATGTGAGATGCACAGAAAGTG	1435
Db	60271	CTACTTCTCTTCCAGTGCATATGCCACCTGGGACCTTATGTGAGATGCACAGAAAGTG	60212
QY	1436	GACACAGGGCCAGGGTTCCTCAAAAAGAAATAGAGCTCTGGGGGGTCTGACTGTAGT	1495
Db	60211	GACACAGGGCCAGGGTTCCTCAAAAAGAAATAGAGCTCTGGGGGGTCTGACTGTAGT	60152
QY	1496	TCTTGAAGTTGGGGTTTCCAGTACCATCTGTAGTCCCTGCTGTAGAGCCCATTTCAAC	1555
Db	60151	TCTTGAAGTTGGGGTTTCCAGTACCATCTGTAGTCCCTGCTGTAGAGCCCATTTCAAC	60092
QY	1556	TCCCCACCATTTAACAGAGGCCCAACACACAGAGTGAACAACACCCTGAGAGTCAACAGGA	1615
Db	60091	TCCCCACCATTTAACAGAGGCCCAACACACAGAGTGAACAACACCCTGAGAGTCAACAGGA	60032
QY	1616	AGTCATTTTCAAAAACTCAAGTCTGTTGAGACCAACACATACCTCAGAAAGTAGG	1675
Db	60031	AGTCATTTTCAAAAACTCAAGTCTGTTGAGACCAACACATACCTCAGAAAGTAGG	59972
QY	1676	ACTGTGGCTTAGAGGGAAGAAAGAAAGCTGATGATGTCTTACCGTACACAGATCTTG	1735
Db	59971	ACTGTGGCTTAGAGGGAAGAAAGAAAGCTGATGATGTCTTACCGTACACAGATCTTG	59912
QY	1736	GATGATCCAGAGCTATATGTACCTCAGAGCAAGAAAGAACTTGGAGAGCTTAGATC	1795
Db	59911	GATGATCCAGAGCTATATGTACCTCAGAGCAAGAAAGAAAGCTTGGAGAGCTTAGATC	59852
QY	1796	CTCAAAATGCCCAATTTAGAGAACACAGCCCACTCTTTTCTTTTGTGAGACGA	1855
Db	59851	CTCAAAATGCCCAATTTAGAGAACACAGCCCACTCTTTTCTTTTGTGAGACGA	59793
QY	1856	GTCTTGCCCTGTGTTCCCATGCTGAGATGCAATGCAATGCATCTCAGCTCACTGCACCTCC	1915
Db	59792	GTCTTGCCCTGTGTTCCCATGCTGAGATGCAATGCAATGCATCTCAGCTCACTGCACCTCC	59733
QY	1916	ATCTCTGATTTCAAAACAATTTCCGGCTCAGGCTCCAGATTAACCTGGATTACAGGGC	1975

775 CCGGAGAGCCCAAGACCTAAGGAAGAAGCCTCTTAATCATCTCTCTTAACTTCCTCTCTTTT  
Db 301 AGGGAGAAACAGAGGAACCTGCTTTGTATTACTGGGAACCCTTGAGAGGGGCCCTTCTCCATC 360

Db	361	CGGAGACCCAGACAGAGAGGGCTTTAACTCTGTGCATGTCGCCGCTCAGCGCCCTGCA	420
Oy	835	TCTTGAGACCGGATCAGACACTACAGAGTCCACTGCTTTGAACAATGCTGGCTGTACATC	894
Db	421	TCTGTGGAGCCGGATCAGACACTACAGAGTCCAAGTCCCTTGAACAATGGCTGGCTGTACATC	480
Oy	895	TCACCGCGCTCATCTTCCCCTCACCTCCAGGACCCTGGTAGCACCATTACTGACCTTGAGG	954
Db	481	TACCGCGCTCATCTTCCCCTCACCTCCAGGACCCTGGTAGCACCATTACTGACCTTGAGG	540
Oy	955	GATGACATCTGTCTGCTACTCAAGAAGCCCTGTGTCTCTGANAAGGGCTGGCCCTTCCTT	1014
Db	541	GATGACATCTGTCTGCTACTCAAGAAGCCCTGTGTCTCTGANAAGGGCTGGCCCTTCCTT	600
Oy	1015	GGCAGAGATATACCCCTACTACTGTGACTGTGAGAGAGACACACTCACTGAGAAAGACTG	1074
Db	601	GGCAGAGATATACCCCTACTACTGTGACTGTGAGAGAGACACACTCACTGAGAAAGACTG	660
Oy	1075	GACAGCTCCCTCTCTGTTTTTTGTGAAGCTGCGACAGGGAGAGAGTCTTCTCAAGAGAGGT	1134
Db	661	GACAGCTCCCTCTCTGTTTTTTGTGAAGCTGCGACAGGGAGAGAGTCTTCTCAAGAGAGGT	720
Oy	1135	CTCCGGAGTCCCTCAGCTTCTACATCAGCCTGATGACAGAGGCTGTCTCTTTGGATGAT	1194
Db	721	CTCCGGAGTCCCTCAGCTTCTACATCAGCCTGATGACAGAGGCTGTCTCTTTGGATGAT	780
Oy	1195	GCTCTAG 1200 	
Db	781	GCCTAG 786 	
RESULT 8			
LOCUS	AFJ26353	786 bp mRNA linear PRI 08-NOV-2001	
DEFINITION	Homo sapiens Src-like adapter protein-2 mRNA, complete cds.		
ACCESSION	AFJ26353		
VERSION	AFJ26353.1	GI:16797891	
KEYWORDS	.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 786)		
AUTHORS	Holland,S.J., Liao,X.C., Mendenhall,M.K., Zhou,X., Pardo,J., Chu,P., Spencer,C., Fu,A.C., Sheng,N., Yu,P., Fall,E., Nagin,A., Shen,M., Yu,S., Chan,E., Wu,X., Li,C., Wolsetschlager,W., Aversa,G., Kolbinger,F., Bennett,M.K., Molinaux,S., Luo,Y., Payan,D.G., Mancebo,H.S.Y. and Wu,J. Functional Cloning of Src-Like Adapter Protein-2 (SLAP-2), a Novel Inhibitor of Antigen Receptor Signaling J. Exp. Med. 194 (9), 1263-1276 (2001)		
TITLE	JOURNAL MEDLINE PUBMED		
JOURNAL	11563592		
PUBMED	11563592		
REFERENCE	2 (bases 1 to 786) Holland,S.J., Mendenhall,M.K., Zhou,X., Spencer,C., Pardo,J., Fu,A.C., Sheng,N., Shen,M., Liao,X., Luo,Y., Payan,D.G., Mancebo,H.S.Y. and Wu,J. Direct Submission Submitted (05-DEC-2000) Rigel Pharmaceutical Inc., 240 East Grand Avenue, South San Francisco CA 94080, USA		
AUTHORS	location/Qualifiers		
TITLE	1..786		
JOURNAL	/organism="Homo sapiens"		
PUBMED	/db_xref="taxon:9606"		
REFERENCE	/chromosome="20"		
SOURCE	1..786		
CDS	/note="SLAP2, SH2 and SH3 domain-containing adaptor" /codon_start=1 /product="Src-like adapter protein-2" /protein_id="AAI29204.1" /db_xref="GI:16797892" /translation="WGSLSRKRLSPPSLSSVGQGPVTMEKSKATVALEFPI AGGEPAISLKGEPILTVSEDDMWITVALSVSGREYNIPSYAVAKYSHGLVEGSR		

BASE COUNT	162 a	234 c	231 g	159 t
ORIGIN	KAEELLLLEPGNGGAGFLIRRESQTRGYSISLAPSLPASMWRIRHRIHCDNGMLVIT SPRTTFPSLQALVDHYSELADICCLKEPVDLQARGPLPDKDILPVPVORTPLNNMK ELDSLLIFSEATGESLISEGIRBSLFTYISLANDAVALDAA"			

Query Match	30.6%;	Score 786;	DB 9;	Length 786;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 786; Conservative	0;	Mismatches	0;	Indels 0;
		Gaps	0;	

QY	415	ATGGAGATCTGGCCAGCAGAAAGAAATCTCTGCAAGCCCAAGCTTGAGTCTCTGTG	474
Db	1	ATGGAGATCTGGCCAGCAGAAAGAAATCTCTGCAAGCCCAAGCTTGAGTCTCTGTG	60
QY	475	CAAGGCCAGGSACTGTGACCACTGGAAACAGAGAAACCAAGCCCAAGCCCTGGCCCTG	534
Db	61	CAAGGCCAGGSACTGTGACCACTGGAAACAGAGAAAGCAAGCCCAAGCCCTGGCCCTG	120
QY	535	GAGCATTTCCGGCAGGTGGCCCGGCCAGCTGTGCTGAGACTTGAGGAGCCATTGACC	594
Db	121	GAGCATTTCCGGCAGGTGGCCCGGCCAGCTGTGCTGAGACTTGAGGAGCCATTGACC	180
QY	595	ATGTGTTCTGAGAGATGAGAACCTGTGTGACCGGTGTGTGAAAGTCTCAGGAGAGAGTAT	654
Db	181	ATGTGTTCTGAGAGATGAGAACCTGTGTGACCGGTGTGTGAAAGTCTCAGGAGAGAGTAT	240
QY	655	AACATCCCAAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATAGAGGCTTGACC	714
Db	241	AACATCCCAAGCGTCCACGTGGCCAAAGTCTCCCATGGGTGGCTGTATAGAGGCTTGACC	300
QY	715	AGGGAAAGAGCAGAGAGAACTGCTGTTGTTTCACTGGAAACCTTGAGAGGGACCTTCTCATC	774
Db	301	AGGGAAAGAGCAGAGAGAACTGCTGTTGTTTCACTGGAAACCTTGAGAGGGACCTTCTCATC	360
QY	775	CGGAGAGCCAGACCAAGAGAGGCTTTACTTCTGTCAGTCCGCTTACGCCCTCTGCA	834
Db	361	CGGAGAGCCAGACCAAGAGAGGCTTTACTTCTGTCAGTCCGCTTACGCCCTCTGCA	420
QY	835	TCTGTGGACCCGATTCAGACACTACAGAGATCCAGCTGCTTGAACAATGCTGCTGTACATC	894
Db	421	TCTGTGGACCCGATTCAGACACTACAGAGATCCAGCTGCTTGAACAATGCTGCTGTACATC	480
QY	895	TCAACCGGCTTGACCTTCCCTCTCACTCCAGGCTCTGTGTGACCATTAATCTGAGACTGGG	954
Db	481	TCAACCGGCTTGACCTTCCCTCTCACTCCAGGCTCTGTGTGACCATTAATCTGAGACTGGG	540
QY	955	GATGATCATGTGCTGCTCACTCAAGAGGCTGTGTCTCTGAGAGGCTTGCCCTCTCCCT	1014
Db	541	GATGATCATGTGCTGCTCACTCAAGAGGCTGTGTCTCTGAGAGGCTTGCCCTCTCCCT	600
QY	1015	GCGCAAGATATATCCCTCACTGATGACTGTGAGAGAGCAACACTCAACTGAGAAAGACCTG	1074
Db	601	GCGCAAGATATATCCCTCACTGATGACTGTGAGAGAGCAACACTCAACTGAGAAAGACCTG	660
QY	1075	GACAGCTCCCTCTGTTTCTGAAAGCTGCAACAGGGAGAGATCTCTTCTCACTGAGAGGT	1134
Db	661	GACAGCTCCCTCTGTTTCTGAAAGCTGCAACAGGGAGAGATCTCTTCTCACTGAGAGGT	720
QY	1135	CTCCGGGAATCCCTCAAGTTCTATCACTCAAGCCGAATGACGAGGCTGTCTCTTGGATGAT	1194
Db	721	CTCCGGGAATCCCTCAAGTTCTATCACTCAAGCCGAATGACGAGGCTGTCTCTTGGATGAT	780
QY	1195	GCCTAG 1200	
Db	781	GCCTAG 786	

RESULT 9

AF290986

LOCUS

DEFINITION Homo sapiens Strc-like adaptor protein-2 splice isoform mRNA,

complete cds.

ACCESSION AF290986

737 bp

mRNA

linear

PRI 22-APR-2002

VERSION	AF290986.1	GI:17351922
KEYWORDS		
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 737)	
AUTHORS	Loreto, M.P., and McClade, C.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (28-JUL-2000) Brain Tumour Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ont M5G 1X8, Canada	
FEATURES	<p>location/Qualifiers</p> <p>1..737</p> <p>    /organism="Homo sapiens"</p> <p>    /db_xref="taxon:9606"</p> <p>    /chromosome="20"</p> <p>    /tissue_type="thymus"</p> <p>    1..633</p> <p>        /note="SLAP-2-v"</p> <p>        /codon_start=1</p> <p>        /product="Src-like adaptor protein-2 splice isoform"</p> <p>        /protein_id="AAB38196.1"</p> <p>        /db_xref="GI:17351923"</p> <p>        /translation="MGSLSPRRSLSPSLSSVGGPVTMEARSKATAVALGSFFAGAGLSTLRLGPEPLTISDEGDMWTVLSPVSGRPNIPYVAVKSHQMLVEGSLRKAELLALLPENPGCAAFILRESOTRGSYSLSVRLSPAPSMPIRHHICLDNGLWLYSPRLFPSTLADVHSSEMPAPMGGTPTPCACADDTTLQLBRAGQLPVPF"</p>	
BASE COUNT	152 a	219 c 218 g 148 t
ORIGIN		
Query Match	20.8%; Score 534; DB 9; Length 737;	
Basic Local Similarity	100.0%; Pred. No. 1.9e-291;	
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	415	ATGGGAAGTCTCCGACAGAGAAGAAATCTCGCAAGCCCAAGCTTGAGTTCCCTGTGC 474
DB	1	ATGGGAAGTCTGCCACAGAGAAGAAATCTCGCAAGCCCAAGCTTGAGTTCCCTGTGC 60
QY	475	CAAGGCCAGGACCTGTGACATGAGAAAGAGAGAAAGCAAGGCCACAGCCGTGGCCCTG 534
DB	61	CAAGGCCAGGACCTGTGACATGAGAAAGAGAGAAAGCAAGGCCACAGCCGTGGCCCTG 120
QY	535	GCGAGTTTCCCGGAGGTGGCCGGCCGAGAGCTGTGCTGAGACCTCGGGAGGCCATTGAC 594
DB	121	GCGAGTTTCCCGGAGGTGGCCGGCCGAGAGCTGTGCTGAGACCTCGGGAGGCCATTGAC 180
QY	595	ATGCTCTGAGAGATGAGACTGGTGGACGGTGTCTGTGAAGTCTCAGGACAGAGATAT 654
DB	181	ATGCTCTGAGAGATGAGACTGGTGGACGGTGTCTGTGAAGTCTCAGGACAGAGATAT 240
QY	655	AACATCCCAAGGCTCAGGTGGCCAAAGTCTCCATGGGTGGCTGATAGAGGCTGAGC 714
DB	241	AACATCCCAAGGCTCAGGTGGCCAAAGTCTCCATGGGTGGCTGATAGAGGCTGAGC 300
QY	715	AGGAGAAAGCAGAGAACTGCTGTGTACTCTGGGAAACCTGTGAGAGGGCTTCTCATC 774
DB	301	AGGAGAAAGCAGAGAACTGCTGTGTGTACTCTGGGAAACCTGTGAGAGGGCTTCTCATC 360
QY	775	CGGAGAGCCAGACCCAGAGAGGCTTTACTTGTTCAGTTCGCTCAGCCGAGCCCTGCA 834
DB	361	CGGAGAGCCAGACCCAGAGAGGCTTTACTTGTTCAGTTCGCTCAGCCGAGCCCTGCA 420
QY	835	TTCGTGGACCGGATCAGACCTTCAGAGATCCACGTGACCTTCAGAAAGCTGCTGTATATC 894
DB	421	TTCGTGGACCGGATCAGACCTTCAGAGATCCACGTGACCTTCAGAAAGCTGCTGTATATC 480
QY	895	TCAACCGGCTCACCCTTCCCTCAGTCCAGGCGCTGTGTGAGACCATTAATCTTGAG 948
DB	481	TCAACCGGCTCACCCTTCCCTCAGTCCAGGCGCTGTGTGAGACCATTAATCTTGAG 534

LOCUS	HS460J8	linear	PRI 23-JUL-2001
DEFINITION	Human DNA sequence from clone RP3-460J8 on chromosome 20q11.21-11.21 Contains the 3' end of the gene for a novel protein similar to N-myc downstream regulated (NDRG1) the 5' end of a gene encoding a novel protein tyrosine kinase, ESRs, STSs and GSSe, complete sequence.		
ACCESSION	AF031662		
VERSION	AF031662.26	GI:9716901	
KEYWORDS	HTG; NDRG1; SH2 domain.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 65741)		
TITLE	Skuce, C.		
JOURNAL	Direct Submission		
COMMENT	Submitted (23-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Aug 7, 2000 this sequence version replaced gi:6425549.		

FEATURES	source	Location/Qualifiers
		1..66741
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="20"
		/map="q11.21-11.23"
		/clone="RP3-460J8"
		/clone_1fb="RPC1-3"
misc_feature		complement(50..544)
		/note="match: GSS: Em:B45150"
gene		complement(240..7128)
		/gene="d9j77B1.1"
mRNA		complement(join(<240..339,6995..7128))
		/gene="d9j77B1.1"
		/product="d460J8.2 (novel protein tyrosine kinase with Src homology 2 (SH2) domain) "
		/note="match: CDNAS: Em:AK025645
		match: ESTs: Em:BG178487"
		/evidence="not_experimental"
CDS		complement(join(<240..339,6995..7085))
		/gene="d9j77B1.1"



/note="Continues in Em:AL050318 as dJ977B1.2"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dJ460J8.2 (novel protein tyrosine kinase with  
src homology 2 (SH2) domain )"  
/protein\_id="CAC44645.1"  
/db\_xref="GI:15020830"  
/translation="MGSLPSRKSLSPSPSLSSVQGGCVPTWEAKRATKVALGSFP  
AGGPALESLRLGEPPLIVS"  
536..565  
repeat\_region  
/note="15 copies 2 mer ag 93% conserved"  
2375..2402  
/note="14 copies 2 mer ta 100% conserved"  
2783..3155  
/note="match: GSS: Em:AQ0807191"  
complement(6983..7482)  
/note="match: GSS: Em:AQ056467"  
complement(6985..7492)  
/note="match: GSS: Em:AQ056478"  
7486..7925  
/note="match: STS: Em:HS427J1S"  
10615..10640  
/note="13 copies 2 mer tc 92% conserved"  
13593..13961  
/note="MER21B repeat: matches 422..785 of consensus"  
14248..14297  
/note="MER21B repeat: matches 374..422 of consensus"  
14594..14614  
/note="MER21B repeat: matches 355..374 of consensus"  
14746..15078  
/note="MER21B repeat: matches 1..355 of consensus"  
complement(17516)  
17519..17686  
/note="match: GSS: Em:AZ067993"  
complement(17521)  
/gene="dJ469A13.3"  
complement(17522)  
/note="match: STS: Em:G19945"  
join(19202..19451,22110..22282,26087..26127,30796..30847)  
/note="match: STS: Em:G23762"  
complement(join(19270..19451,20551..20589,22110..22161,  
22247..22282,26087..26134,30796..30847,30986..31051,  
32039..32142,37091..37147,46566..46652,48268..48328,  
50163..50225,53242..53362,54429..54534))  
/gene="dJ469A13.3"  
/product="dJ460J8.1 (continued from dJ469A13.3 in  
Em:AL132768)"  
/note="match: CDNAS: Em:AB033922 Em:U52073 Em:DB7953  
Em:X92845 Em:AF004162 Em:U60593 Em:AF045564 Em:AB033921  
Em:AF159092 Em:AF147402 Em:W59814  
match: ESTs: Em:AA718726 Em:AA039000 Em:AM003952 Em:W89263  
Em:AA325826 Em:AI230962 Em:AA162360 Em:AA445016  
Em:AV002395 Em:T85147 Em:AI786615 Em:AA113437 Em:AI004026  
Em:T88705 Em:AI786673 Em:AI181197 Em:AA764653 Em:AA316771  
Em:AU035165 Em:AV002368"  
/evidence=not\_experimental  
complement(17542..17547)  
/gene="dJ469A13.3"  
17555..17811  
/note="match: STS: Em:G19945"  
join(19202..19451,22110..22282,26087..26127,30796..30847)  
/note="match: STS: Em:G23762"  
complement(join(19270..19451,20551..20589,22110..22161,  
22247..22282,26087..26134,30796..30847,30986..31051,  
32039..32142,37091..37147,46566..46652,48268..48328,  
50163..50225,53242..53362,54429..54534))  
/gene="dJ469A13.3"  
/note="novel protein (FLJ13556) similar to N-myc  
downstream regulated (NDRG1)  
match: proteins: Sw:Q62433 Sw:Q92597 Tr:Q9Z2U9 Sw:P97862"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dJ460J8.1 (continued from dJ469A13.3 in  
Em:AL132768)"  
/protein\_id="CAB65625.1"

/db\_xref="GI:6687781"  
/db\_xref="SWISS-PROT:Q9UGV2"  
/translation="EHDIETTHGVAVITIGLPRKGNRPVILTYHDIGLNKSCFNAP  
NEEDMOETHTFAVCHVDAPGOGAPSPFGVQVPTMDLAEMLPVLTNLSKSI  
GIGGAGAYLISRFALNHPELVGLIINDDPCAKMTDMAASLGLSTNNVDITLA  
HRRGFEQANLIDLIOTYRMAADINDODLQFLANSNGRDLIEIRPLIGKNDK  
KTLKSTLIVDNGSPAYAVEGCSRLNINTTLKPADGGGLPVVQGLTEAFK  
YFQMGVITPVVLSHTSEVSPASWTRLSRTHSTSSLSGSPFSRSVTNOS  
DQDQSCSPVLDRIHQTEVSC"  
complement(24246..24780)  
/gene="dJ469A13.3"  
/note="match: GSS: Em:AQ0592789"  
complement(27980..28556)  
/gene="dJ469A13.3"  
/note="match: GSS: Em:AQ080867"  
34313..34420  
/note="HY1 repeat: matches 1..109 of consensus"  
34505..34560  
/note="28 copies 2 mer ta 78% conserved"  
34564..34611  
/note="24 copies 2 mer ta 79% conserved"  
42553..43658  
/note="MER52C repeat: matches 1..1278 of consensus"  
47985..48470  
/note="match: GSS: Em:AQ672417"  
49391..49480  
/note="45 copies 2 mer ta 76% conserved"  
52922..53120  
/note="match: STS: Em:G04621"  
complement(53426..54032)  
/gene="dJ469A13.3"  
/note="match: GSS: Em:AQ314824  
match: STS: Em:G55853"  
54054..54477  
/note="match: GSS: Em:AQ310681"  
57509..57838  
/note="match: STS: Em:G07504"  
complement(57696..58013)  
/note="match: STS: Em:G07632 Em:G07634"  
58856..59001  
/note="73 copies 2 mer at 83% conserved"  
complement(58930..59281)  
/note="match: GSS: Em:AQ067563"  
59062..59236  
/note="TIGER2 repeat: matches 2541..2714 of consensus"  
59738..60495  
/note="TIGER2 repeat: matches 1780..2541 of consensus"  
60719..61831  
/note="MER11C repeat: matches 1..1071 of consensus"  
62547..63174  
/note="TIGER2 repeat: matches 1165..1784 of consensus"  
63475..63703  
/note="TIGER2 repeat: matches 125..354 of consensus"  
64017..64063  
Query Match 14.0% Score 359; DB 9; Length 66741;  
Best Local Similarity 100.0%; Pred.No. 3e-192;  
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 13 GTTCGAGCTAGAGCTCCAAAGACCCACGCGTGTCTCTGTGACAGACTCAAGAGGCC 72  
Db 11945 GTTCGAGCTAGAGCTCCAAAGACCCACGCGTGTCTCTGTGACAGACTCAAGAGGCC 11886  
QY 73 CTGGGCTTCCCTCCCTGCGCGTGGCTGCTGGAGAGGTTCCCACTGACAAATCCCTTA 132  
Db 11886 CTGGGCTTCCCTCCCTGCGCGTGGCTGCTGGAGAGGTTCCCACTGACAAATCCCTTA 11826  
QY 133 AGAGAGCTGGGAGAGCTATCCCTGCTGATGTATCAAACTGCTGACGACAGATGCT 192  
Db 11825 AGAGAGCTGGGAGAGCTATCCCTGCTGATGTATCAAACTGCTGACGACAGATGCT 11766  
QY 199 GAGCTACCAACCAACGCTAGCGTCTCCGAGATCTCCCAAGCGTGGAGAGTCTT 252

Db 11765 GAGCTACCAACCAACCACTAGCTCTCTCTGAAATCTCTCCAGCTGAGAGATTCT 11706  
QY 253 GGGTGTCTTAGGACCAAGACACTGGCAGACTTCCAGAGGGCCCCCAAGCCCTTAACCT 312  
Db 11705 GGGTGTCTTAGGACCAAGACACTGGCAGACTTCCAGAGGGCCCCCAAGCCCTTAACCT 11646  
QY 313 GTCCAGCAGACAGATGCTCTCAGCAGACTGCTCTCCAGAGCCCTTGTATGACAAACCA 371  
Db 11645 GTCCAGCAGACAGATGCTCTCAGCAGACTGCTCTCCAGAGCCCTTGTATGACAAACCA 11587

## RESULT 11

AC026539 145833 bp DNA linear HTG 27-APR-2000  
LOCUS Homo sapiens chromosome 20 clone RP11-712N14 map 20, WORKING DRAFT  
DEFINITION SEQUENCE, 37 unordered pieces.  
AC026539.2 GI:7656813  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 145833)  
AUTHORS Birren, B., Linton, J., Nussbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 20, clone RP11-712N14  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145833)  
AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Baetien, V., Bedal, F.,  
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., S.,  
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galgan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Harford, A., Horton, L.,  
Howard, J., C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,  
Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehoczeky, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, C., McPherson, R.,  
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C., H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Olivari, T., M., Oliver, J., Peterson, K., Pierre, N.,  
Pisanti, C., Pollari, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,  
Theriot, S., Theodore, J., Titrill, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
JOURNAL Submitted, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Apr 27, 2000 this sequence version replaced g1:7283243.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7115

Center clone name: 712 N14

----- Summary Statistics

Sequencing vector: M13, M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 125573 bases at least Q40

Consensus quality: 135703 bases at least Q30

Insert size: 18200; agarose-fp

Insert size: 142233; sum-of-contigs

Quality coverage: 2.6 in Q20 bases; agarose-fp  
Quality coverage: 3.3 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1214: contig of 1214 bp in length  
\* 1215 1314: gap of 100 bp  
\* 1315 2673: contig of 1359 bp in length  
\* 2674 2773: gap of 100 bp  
\* 2774 4520: contig of 1747 bp in length  
\* 4521 4620: gap of 100 bp  
\* 4621 5961: contig of 1341 bp in length  
\* 5962 6061: gap of 100 bp  
\* 6062 7719: contig of 1658 bp in length  
\* 7720 7819: gap of 100 bp  
\* 7820 9799: contig of 1980 bp in length  
\* 9800 9899: gap of 100 bp  
\* 9900 11434: contig of 1535 bp in length  
\* 11435 11534: gap of 100 bp  
\* 11535 14382: contig of 2848 bp in length  
\* 14383 14482: gap of 100 bp  
\* 14483 17148: contig of 2666 bp in length  
\* 17149 17248: gap of 100 bp  
\* 17249 19239: contig of 1991 bp in length  
\* 19240 19339: gap of 100 bp  
\* 19340 21102: contig of 1763 bp in length  
\* 21103 21202: gap of 100 bp  
\* 21203 23371: contig of 2169 bp in length  
\* 23372 23471: gap of 100 bp  
\* 23472 25782: contig of 2311 bp in length  
\* 25783 25882: gap of 100 bp  
\* 25883 28824: contig of 2942 bp in length  
\* 28825 28924: gap of 100 bp  
\* 28925 31619: contig of 2635 bp in length  
\* 31620 31719: gap of 100 bp  
\* 31720 34679: contig of 2960 bp in length  
\* 34680 34779: gap of 100 bp  
\* 34780 37527: contig of 2748 bp in length  
\* 37528 37627: gap of 100 bp  
\* 37628 40146: contig of 2519 bp in length  
\* 40147 40246: gap of 100 bp  
\* 40247 43743: contig of 3497 bp in length  
\* 43744 43843: gap of 100 bp  
\* 43844 47133: contig of 3290 bp in length  
\* 47134 47233: gap of 100 bp  
\* 47234 51023: contig of 3790 bp in length  
\* 51024 51123: gap of 100 bp  
\* 51124 54935: contig of 3812 bp in length  
\* 54936 55035: gap of 100 bp  
\* 55036 59553: contig of 4518 bp in length  
\* 59554 59653: gap of 100 bp  
\* 59654 64428: contig of 4775 bp in length  
\* 64429 64528: gap of 100 bp  
\* 64529 69211: contig of 4683 bp in length  
\* 69212 69311: gap of 100 bp  
\* 69312 72901: contig of 3590 bp in length  
\* 72902 73001: gap of 100 bp  
\* 73002 76724: contig of 3723 bp in length  
\* 76725 76824: gap of 100 bp  
\* 76825 81179: contig of 4355 bp in length  
\* 81180 81279: gap of 100 bp  
\* 81280 87009: contig of 5730 bp in length  
\* 87010 87109: gap of 100 bp  
\* 87110 90855: contig of 3746 bp in length  
\* 90856 90955: gap of 100 bp  
\* 90956 96520: contig of 5565 bp in length  
\* 96521 96620: gap of 100 bp

```

* 96621 102321: contig of 5701 bp in length
* 102322 102421: gap of 100 bp
* 102422 108293: contig of 5872 bp in length
* 108294 108393: gap of 100 bp
* 108394 116689: contig of 8296 bp in length
* 116790 116789: gap of 100 bp
* 116790 125264: contig of 8475 bp in length
* 125265 125364: gap of 100 bp
* 125365 136354: contig of 10990 bp in length
* 136355 136454: gap of 100 bp
* 136455 145833: contig of 9379 bp in length.

```

## FEATURES

```

source
1. 145833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20"
/clone_id="RP11-712N14"
/clone_1b="RPCT-11 Human Male BAC"
1. 1214
/misc_feature
/feature="assembly_fragment"
1315. 2673
/feature="assembly_fragment"
2774. 4520
/feature="assembly_fragment"
4621. 5961
/feature="assembly_fragment"
6062. 7719
/feature="assembly_fragment"
7820. 9799
/feature="assembly_fragment"
9900. 11434
/feature="assembly_fragment"
11535. 14382
/feature="assembly_fragment"
14483. 17148
/feature="assembly_fragment"
17249. 19239
/feature="assembly_fragment"
19340. 21102
/feature="assembly_fragment"
21203. 23371
/feature="assembly_fragment"
23472. 25782
/feature="assembly_fragment"
25883. 28824
/feature="assembly_fragment"
28925. 31619
/feature="assembly_fragment"
31720. 34679
/feature="assembly_fragment"
34780. 37527
/feature="assembly_fragment"
37628. 40146
/feature="assembly_fragment"
40247. 43743
/feature="assembly_fragment"
43844. 47133
/feature="assembly_fragment"
47234. 51023
/feature="assembly_fragment"
51124. 54935
/feature="assembly_fragment"
vector_end:right"
55036. 59553
/feature="assembly_fragment"

```

## Query Match

Best Local Similarity 100.0%; Score 359; DB 2; Length 145833;

Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GTCCAGCTAGAGCTCAAGAGCCCGCTGTGTCTGTGACAGAGCTCAAGAGGCC 72

```

Db 99740 GTCCAGCTAGAGCTCAAGAGCCCGCTGTGTCTGTGACAGAGCTCAAGAGGCC 99799
QY 73 CTGGAGCTTCTCCCTCCCTGAGCTCGAGCTGTGTGGAGGGTTCCCACTCAGATCCCTTA 132
Db 99800 CTGGAGCTTCTCCCTCCCTGAGCTCGAGCTGTGTGGAGGGTTCCCACTCAGATCCCTTA 99859
QY 133 AGGAGCATGGGAGAGCTATTCATCCCTGTGTGAACAACTCTGACAGACATCTCT 192
Db 99860 AGGAGCATGGGAGAGCTATTCATCCCTGTGTGAACAACTCTGACAGACATCTCT 99919
QY 193 GAGCTACCCAAACCAACCTAGCTCTCCCTGAAATCTCCAGGCTGAGAGATTCT 252
Db 99920 GAGCTACCCAAACCAACCTAGCTCTCCCTGAAATCTCCAGGCTGAGAGATTCT 99979
QY 253 GGGGTGCTCTGAGACCAAGAGACCTGGAGACCTTCCAGAGGGCCCCCAAGCCCTAACCT 312
Db 99980 GGGGTGCTCTGAGACCAAGAGACCTGGAGACCTTCCAGAGGGCCCCCAAGCCCTAACCT 100039
QY 313 GTCCAGCTAGAGCTCGCTTCAGACAGAGCTGTCTTCCCAAGCCTTGTGATGACAAACCA 371
Db 10040 GTCCAGCTAGAGCTCGCTTCAGACAGAGCTGTCTTCCCAAGCCTTGTGATGACAAACCA 100098

```

```

RESULT 12
AL353754/c 173316 bp DNA linear PRI 30-JAN-2002
LOCUS Human DNA sequence from clone RP11-401B9 on chromosome 10, complete
DEFINITION
ACCESSION AL353754
VERSION AL353754
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 173316)
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:159623548.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-401B9 is from the library RPCT-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.choi.org/bacpac/home.htm
VECTOR: pBRc3.6

```

## COMMENT

This sequence is the entire insert of clone RP11-401B9 The true left end of clone RP11-554P1 is at 118361 in this sequence. The true right end of clone RP11-29903 is at 102633 in this sequence.

## FEATURES

Location/Qualifiers

```

source
1..173316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-401B3"
/clone_1b="RP11-11.2"
/complement(65129..65132)
/notice="119 bases of 1S5 (J01735) removed here. This
sequence represents the duplicated flanking sequence of
the ISS."

misc_feature
Query Match 2.9%; Score 74; DB 9; Length 173316;
Best Local Similarity 100.0%; Pred. No. 2,2e-30;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2035 TTGGCAGAGCGTGTGCAATCTCGAATCACTCAGAGATGCACCACTTGCCCTCCCAA 2094
|||||
Db 109183 TTGGCAGAGCTGTGTGCAATCTCTAATCTCAGATGATCCACCACTTGCCCTCCCAA 109124
|||||

Oy 2095 GTGCTGGGATTACA 2108
|||||
Db 109123 GTGCTGGGATTACA 109110
|||||

RESULT 13
AL157361/c
LOCUS AL157361 151603 bp DNA linear PRI 24-MAY-20000
DEFINITION Human DNA sequence from clone RP11-132L12 on chromosome
13q21.33-22.3 Contains an STS and GSS, complete sequence.
ACCESSION AL157361
VERSION AL157361.6 GI:7899161
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 151603)
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (24-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT
Requester: clonequest@sanger.ac.uk
On May 17, 2000 this sequence version replaced gi:7739119.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
feature is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-132L12 is from the library RP11-11.1 constructed at the
Rowell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://baepc.med.buffalo.edu/
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-132L12. The true
left end of clone RP11-13502 is at 131966 in this sequence.
Location/Qualifiers

```

```

repeat_region 23502..23547
/note="23 copies 2 mer at 78% conserved"
repeat_region 24310..24754
/note="L2 repeat: matches 2275..2749 of consensus"
repeat_region 25348..25561
/note="MIR repeat: matches 8..224 of consensus"
misc_feature 25847..26252
/note="match: GSS: Em:AQ602207"
repeat_region 25911..25955
/note="MIR repeat: matches 99..145 of consensus"
repeat_region 27540..27585
/note="MIR repeat: matches 216..261 of consensus"
repeat_region 28853..29156
/note="ALUX repeat: matches 1..309 of consensus"
repeat_region 31217..31248
/note="16 copies 2 mer ac 96% conserved"
repeat_region 31397..31608
/note="HAL1 repeat: matches 1191..1409 of consensus"
repeat_region 31670..31963
/note="ALUX repeat: matches 1..291 of consensus"
repeat_region 32072..32223
/note="MSTC repeat: matches 1..152 of consensus"
repeat_region 32315..32428
/note="MSTC repeat: matches 248..363 of consensus"
repeat_region 32429..32664
/note="LIMB7 repeat: matches 309..553 of consensus"
repeat_region 32668..34742
/note="LIP3 repeat: matches 4075..6146 of consensus"
repeat_region 34775..35227
/note="LIMB7 repeat: matches 538..970 of consensus"
repeat_region 35228..35531
/note="ALUB repeat: matches 1..301 of consensus"
repeat_region 35532..36720
/note="LIMB7 repeat: matches 970..2189 of consensus"
repeat_region 36886..37021
/note="LIMB7 repeat: matches 2204..2341 of consensus"
repeat_region 37029..37220
/note="ALUB repeat: matches 107..304 of consensus"
repeat_region 37230..37345
/note="58 copies 2 mer ga 74% conserved"
repeat_region 37347..38783
/note="LIMB7 repeat: matches 2380..3851 of consensus"
repeat_region 38787..39989
/note="LIMB7 repeat: matches 3996..5195 of consensus"
repeat_region 39996..40110
/note="FLAM C repeat: matches 7..121 of consensus"
repeat_region 40111..40144
/note="BC200 repeat: matches 7..158 of consensus"
repeat_region 40145..40300
/note="LIMB7 repeat: matches 5200..5366 of consensus"
repeat_region 40304..40918
/note="LIMB7 repeat: matches 5527..6165 of consensus"
misc_feature 41788..42279
/note="match: GSS: Em:AQ587336"
repeat_region 42229..42641
/note="MER65A repeat: matches 1..443 of consensus"
repeat_region 43403..43497
/note="MIR repeat: matches 84..190 of consensus"
repeat_region 43514..45711
/note="MIR repeat: matches 3..199 of consensus"
repeat_region 46007..46019
/note="MIR repeat: matches 16..28 of consensus"
repeat_region 46020..46313
/note="ALUSG repeat: matches 1..296 of consensus"
repeat_region 46314..46522
/note="MIR repeat: matches 28..228 of consensus"
repeat_region 46610..46822
/note="ALUB repeat: matches 14..226 of consensus"
repeat_region 46823..47188
/note="THELB repeat: matches 1..371 of consensus"
repeat_region 47189..47274
/note="ALUB repeat: matches 226..311 of consensus"
repeat_region 48568..48726

```

```

repeat_region 49609..49812
/note="ALUB/FRAM repeat: matches 157..304 of consensus"
repeat_region 49813..50168
/note="LIMC4 repeat: matches 6635..6820 of consensus"
repeat_region 50169..50306
/note="THELB repeat: matches 1..364 of consensus"
repeat_region 50307..50892
/note="LIMC4 repeat: matches 6820..7007 of consensus"
/note="LIMC4 repeat: matches 7168..7780 of consensus"

Query Match 2.8%; Score 73; DB 9; Length 151603;
Best Local Similarity 100.0%; Pred. No. 8.4e-29;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCGACCTGAGTGATCCACGACCTGCGCTCCCAAGTCGGGATTACAG 2109
Db 69231 TCGAAGCTCGACCTGAGTGATCCACGACCTGCGCTCCCAAGTCGGGATTACAG 69172
QY 2110 GTGTGAGCCACGG 2122
Db 69171 GTGTGAGCCACGG 69159

RESULT 14
AC092450_1
WPCOMMENT
Sequence split into 8 fragments
Fragment Name Begin End
AC092450_0 1 110000
AC092450_1 100001 210000
AC092450_2 200001 310000
AC092450_3 300001 410000
AC092450_4 400001 510000
AC092450_5 500001 610000
AC092450_6 600001 710000
AC092450_7 700001 727300
Continuation (2 of 8) of AC092450 from base 100001 (AC092450 Homo sapiens chromosome 12

Query Match 2.8%; Score 72; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 3.2e-29;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2039 CCAGGCTGGTGTGAACTCTGACCTGAGTGATCCACGACCTTGCTCCCAAGTGC 2098
Db 97888 CCAGGCTGGTGTGAACTCTGACCTGAGTGATCCACGACCTTGCTCCCAAGTGC 97947
QY 2099 TGGGATTACAGG 2110
Db 97948 TGGGATTACAGG 97959

RESULT 15
AL356532/c
LOCUS
DEFINITION
Human DNA sequence from clone RP3-503M14 on chromosome 1. Contains
Esrs, GSSs and Cpg islands. Contains the 5' part of the MP2L1 gene
for myelin protein zero-like 1, complete sequence.
ACCESSION
AL356532
VERSION
AL356532.9 GI:10120062
KEYWORDS
HTG; Cpg island; MP2L1; myelin protein.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 24706)
AUTHORS
Coville,G.
TITILE
Direct Submission
JOURNAL
Submitted (19-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
On Sep 13, 2000 this sequence version replaced gi:9931783.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations

```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/Celegans/wormpep>. This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

IMPORTANT: This sequence is not the entire insert of clone RP3-503M14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP1-313L4 is at 24607 in this sequence. The true right end of clone RP3-455U7 is at 100 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

RP3-503M14 is from the library RPCT-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

#### FEATURES

##### source

```

1. 24706
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP3-503M14"
/clone_lib="RPCT-3"
/complement(1..313)
/note="match: GSS: Em:AQ405985"
complement(1..228)
/note="match: GSS: Em:AQ523677"
12..76
/note="L2 repeat: matches 2074..2231 of consensus"
143..602
/note="MLTIC repeat: matches 34..485 of consensus"
607..1026
/note="L2 repeat: matches 2274..2750 of consensus"
1271..1587
/note="AluY repeat: matches 2..311 of consensus"
2156..2195
/note="AluY repeat: matches 2..311 of consensus"
2156..2195
/note="10 copies 4 mer ttc 100% conserved"
2197..2508
/note="AluX repeat: matches 1..312 of consensus"
2851..3033
/note="L2 repeat: matches 2662..2747 of consensus"
3402..3694
/note="AluX repeat: matches 3..299 of consensus"
4275..4392
/note="FLAM_C repeat: matches 1..118 of consensus"
5257..5431
/note="MIR repeat: matches 47..235 of consensus"
5956..6003
/note="3 copies 16 mer 93% conserved"
7175..7234
/note="15 copies 4 mer ctat 71% conserved"
8555..8895
/note="AluX repeat: matches 1..312 of consensus"
9523..9602
/note="L2 repeat: matches 2618..2708 of consensus"
10000..10340
/note="match: GSS: Em:AQ705822"
10288..10585
/note="AluX repeat: matches 1..305 of consensus"
10673..10983
/note="AluX repeat: matches 1..311 of consensus"
complement(11466..11964)
misc_feature

```

```

/note="match: GSS: Em:AQ322449"
12033..12513
/note="match: GSS: Em:AQ169552"
12918..13239
/note="AluY repeat: matches 1..302 of consensus"
13732..14028
/note="AluSP repeat: matches 1..308 of consensus"
14906..15055
/note="MIR repeat: matches 67..222 of consensus"
15144..15694
/note="CpG island"
/evidence="not_experimental"
15628..15699
/note="tRNA-Pro-CCG repeat: matches 1..72 of consensus"
16391..16462
/note="tRNA-Pro-CCT repeat: matches 1..72 of consensus"
17210..17514
/note="AluX repeat: matches 1..309 of consensus"
17526..17827
/note="AluY repeat: matches 1..311 of consensus"
complement(17550..18089)
/note="match: GSS: Em:AQ427854"
complement(17632..18089)
/note="match: GSS: Em:AQ886051"
complement(17677..18089)
/note="match: GSS: Em:AQ423300"
17942..18013
/note="AluY/FRAM repeat: matches 221..292 of consensus"
18026..18556
/note="MER41D repeat: matches 1..557 of consensus"
18683..18718
/note="9 copies 4 mer tcat 86% conserved"
18725..19011
/note="AluS9 repeat: matches 1..299 of consensus"
20479..20540
/note="2 copies 31 mer 93% conserved"
20722..20833
/note="AluS9 repeat: matches 200..309 of consensus"
20834..21181
/note="LTR2 repeat: matches 105..449 of consensus"
21304..21508
/note="AluS9 repeat: matches 1..203 of consensus"
complement(21397..21865)
/note="match: GSS: Em:AQ033633"
22645..23518
/note="CpG island"
/evidence="not_experimental"
22896..23145
/note="MPZL1"
22896..23145
/gene="MPZL1"
22896..23145
/gene="MPZL1"
/product="d0503M14.1 (myelin protein zero-like 1)"
/note="match: cDNAs: Em:AF087020 Em:AK000286 Em:AF092425 Em:AF181660 Em:AF095726 Em:Y17043 Em:AF095727"
/evidence="not_experimental"
23055..23145
/gene="MPZL1"
/note="continues in d31314 in d31314.SPAN 2"
match: proteins: Sw:P03181 Sw:Q92519 Tr:O41980 Tr:O69118 Tr:Q9X4P5 Tr:Q9Q6J3"
/codon_start=1
/evidence="not_experimental"
/product="d0503M14.1 (myelin protein zero-like 1)"
/protein_id="CAC36033.1"
/db_xref="GI:13559053"
/translation="MAASAGAVIAAPSRRLWVLAALGL"
BASE COUNT 6599 a 5837 c 5455 g 6815 t
ORIGIN
Query Match 2.8%; Score 71; DB 9; Length 24706;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

repeat_region /rpt_family="LINE2"
complement(3219..3430)
/rpt_family="LINE2"
repeat_region complement(3907..4016)
/rpt_family="(GGA)n"
repeat_region 4115..4273
/rpt_family="LINE2"
repeat_region complement(4356..4646)
/rpt_family="AluJo"
repeat_region complement(5377..5676)
/rpt_family="AluY"
repeat_region 5861..6011
/rpt_family="LINE2"
complement(6042..6126)
/rpt_family="MER5B"
repeat_region 6179..6545
/rpt_family="LINE2"
repeat_region 6779..7078
/rpt_family="AluSx"
repeat_region 7775..8156
/rpt_family="LINE2"
complement(8685..8765)
/rpt_family="MTR"
repeat_region complement(10030..10749)
/rpt_family="LINE2"
repeat_region 10887..10976
/rpt_family="LINE2"
repeat_region 11106..11402
/rpt_family="AluJo"
complement(11730..11805)
/rpt_family="LINE2"
repeat_region 12330..13004
/rpt_family="LINE2"
repeat_region 13043..13188
/rpt_family="MLT1A2"
complement(13189..13490)
/rpt_family="AluSp"
repeat_region 13528..13678
/rpt_family="MLT1A2"
complement(13692..13740)
/rpt_family="L1MB7"
repeat_region 13742..14034
/rpt_family="AluSc"
complement(14035..14130)
/rpt_family="L1MB8"
complement(14130..14469)
/rpt_family="L1MB8"
complement(14480..14781)
/rpt_family="AluJo"
repeat_region 14803..15004
/rpt_family="LINE2"
complement(15068..15155)
/rpt_family="L1PA15"
complement(15162..15445)
/rpt_family="AluSg"
complement(15639..15724)
/rpt_family="(TA)n"
complement(15725..16021)
/rpt_family="AluSg"
complement(16145..16535)
/rpt_family="MSTD"
repeat_region 16592..16832
/rpt_family="AluY"
repeat_region 16833..16976
/rpt_family="AluSg/x"
repeat_region 16991..17325
/rpt_family="L1MC1"
repeat_region 17330..17641
/rpt_family="AluJo"
repeat_region 17670..17835
/rpt_family="FRAM"
complement(17924..18232)
/rpt_family="AluSx"

```

```

repeat_region 18243..18771
/rpt_family="LINE2"
complement(18325..18444)
/rpt_family="LINE2"
repeat_region 18797..19097
/rpt_family="AluSx"
complement(19138..19370)
/rpt_family="MLT1C"
repeat_region 19399..19610
/rpt_family="AluSg"
complement(19618..19804)
/rpt_family="MLT1C"
complement(19863..19980)
/rpt_family="FLAM C"
complement(20138..20477)
/rpt_family="L1MC4"
repeat_region 20478..20782
/rpt_family="AluSg"
repeat_region 20804..21109
/rpt_family="AluSx"
complement(21224..21337)
/rpt_family="AluJo"
complement(21768..22315)
/rpt_family="MER51B"
repeat_region 22391..22686
/rpt_family="AluSx"
complement(22695..22891)
/rpt_family="L1MC/D"
complement(22927..23211)
/rpt_family="AluSx"
repeat_region 23383..23497
/rpt_family="MIR"
complement(23633..23937)
/rpt_family="AluSx"
complement(24057..24291)
/rpt_family="LINE2"
repeat_region 24536..24645
/rpt_family="(TA)n"
complement(24687..24926)
/rpt_family="AluJo"
repeat_region 25366..25764
/rpt_family="AluSg"
complement(25907..26007)
/rpt_family="MIR"

```

Query Match 2.8% Score 71: DB 9: Length 38875;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-28; Indels 0; Gaps 0;  
 Matches 71; Conservative 0; Mismatches 0;

```

QY 2050 TCGAATCTCTGATCGATGATCCACCCAGCTTGGCTCCCAAGTCTGGATTACAG 2109
DB 13409 TCGAATCTCTGATCGATGATCCACCCAGCTTGGCTCCCAAGTCTGGATTACAG 13468
QY 2110 GTGTGAGCCAC 2120
DB 13469 GTGTGAGCCAC 13479

```

RESULT 18  
 AL596188/c 69676 bp DNA linear PRI 13-SEP-2001  
 LOCUS Human DNA sequence from clone CTA-3109 on chromosome 6, complete  
 DEFINITION sequence.  
 ACCESSION AL596188  
 VERSION AL596188.6 GI:15620698  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.



REFERENCE 1 (bases 1 to 69676)  
 AUTHORS Tracey, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk

## COMMENT

On Sep 14, 2001 this sequence version replaced gi:15591680.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL, Sw:  
 SWISSPROT, Tr: TrEMBL, Wp: WormPEP, Information on the WormPEP  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>  
 CTA-31u9 is from the human BAC library described in U-J. Kim et al.  
 (1996) Genomics 34, 213-218.  
 VECTOR: PBAC108L.  
 IMPORTANT: This sequence is not the entire insert of clone CTA-31u9  
 It may be shorter because we sequence overlapping sections only  
 once, except for a short overlap.  
 The true right end of clone RP11-57H15 is at 2000 in this  
 sequence.

## FEATURES

source 1..69676  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="CTA-31u9"  
 /clone\_11b="CTT978SK-A1"  
 18807..19234  
 /note="Cpg island"  
 /evidence=not experimental  
 misc\_feature 60040..60136  
 /note="Single clone region. Sequence from reads from a  
 short insert library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."  
 BASE COUNT 21399 a 13905 c 14237 g 20135 t  
 ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 69676;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCTTGACCTGAGTATCCACCACCTTGCCCTCCCAAGGCTGGGATTACAG 2109  
 Db 56824 TCGAATCTTGACCTGAGTATCCACCACCTTGCCCTCCCAAGGCTGGGATTACAG 56765  
 QY 2110 GTGTGAGCCAC 2120  
 Db 56764 GTGTGAGCCAC 56754  
 RESULT 19  
 AL445187/c 76114 bp DNA linear PRI 05-DEC-2001  
 LOCUS Human DNA sequence from clone RP11-576C12 on chromosome 9 Contains  
 DEFINITION a VP83JB pseudogene, complete sequence.  
 ACCESSION AL445187

VERSION AL445187.7 GI:11544549  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Submitted (05-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Dec 4, 2000 this sequence version replaced gi:11340318.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL, Sw: SWISSPROT, Tr: TrEMBL, Wp: WormPEP, Information  
 on the WormPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr9>  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. RP11-576C12 is from  
 the library RP11-11.2 constructed by the group of Pieter de Jong.  
 For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: PBAC3.6

## COMMENT

IMPORTANT: This sequence is not the entire insert of clone  
 RP11-576C12 it may be shorter because we sequence overlapping  
 sections only once, except for a short overlap.  
 The true right end of clone RP11-576C12 is at 76114 in this  
 sequence. The true left end of clone RP11-276E15 is at 11501 in  
 this sequence. The true right end of clone RP11-32M23 is at 100 in  
 this sequence.

## FEATURES

source 1..76114  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-576C12"  
 /clone\_11b="RP11-11.2"  
 1..232  
 repeat\_region 1049..1244  
 /note="LM4 repeat: matches 3850..4086 of consensus"  
 repeat\_region 1478..1607  
 /note="MIR repeat: matches 19..256 of consensus"  
 repeat\_region 2107..2406  
 /note="AluY repeat: matches 1..133 of consensus"  
 repeat\_region 2696..2818  
 /note="AluY repeat: matches 1..300 of consensus"  
 repeat\_region 3445..3700  
 /note="AluJ/FRAM repeat: matches 153..289 of consensus"  
 repeat\_region 3881..4168  
 /note="MIR1J repeat: matches 231..488 of consensus"  
 repeat\_region 4805..5110  
 /note="AluY repeat: matches 1..306 of consensus"  
 repeat\_region 5126..5171  
 /note="AluY repeat: matches 1..301 of consensus"  
 repeat\_region 5451..5768  
 /note="23 copies 2 mer aa 87% conserved"  
 misc\_feature /note="match. GSS: Em:AQ018476"

```

misc_feature      complement(5467..5821)
                   /note="match: GSS: Em:AQ128048"
repeat_region     5825..6127
                   /note="AluSq repeat: matches 1..301 of consensus"
misc_feature      complement(6173..6510)
                   /note="match: GSS: Em:AQ347260"
misc_feature      complement(6173..6457)
                   /note="match: GSS: Em:AQ698025"
misc_feature      6466..7152
                   /note="match: GSS: Em:AQ350137"
repeat_region     6892..7434
                   /note="MLT2D repeat: matches 9..553 of consensus"
repeat_region     7438..7556
                   /note="HERVL repeat: matches 5643..5761 of consensus"
repeat_region     7665..7834
                   /note="MLT1J repeat: matches 1..180 of consensus"
repeat_region     7877..7999
                   /note="FLAM_C repeat: matches 5..130 of consensus"
repeat_region     8001..8086
                   /note="43 copies 2 mer aa 76% conserved"
repeat_region     8943..9042
                   /note="L2 repeat: matches 2214..2313 of consensus"
misc_feature      9154..9172
                   /note="Single clone region. Assembly confirmed by
                   restriction digest data."
misc_feature      9173..9245
                   /note="Single clone region. Assembly confirmed by
                   restriction digest data."
misc_feature      complement(11049..11541)
                   /note="match: GSS: Em:AQ678143"
repeat_region     12551..13531
                   /note="LIMB7 repeat: matches 5155..6173 of consensus"
misc_feature      complement(13184..13755)
                   /note="match: GSS: Em:B47740"
misc_feature      13772..14259
                   /note="match: GSS: Em:AQ608809"
repeat_region     15365..15813
                   /note="LIMB7 repeat: matches 4783..5233 of consensus"
repeat_region     16572..16647
                   /note="38 copies 2 mer at 72% conserved"
repeat_region     16711..17016
                   /note="AluSx repeat: matches 3..308 of consensus"
repeat_region     17295..17505
                   /note="MIR repeat: matches 34..250 of consensus"
repeat_region     17626..17696
                   /note="MER5A repeat: matches 12..79 of consensus"
gene              complement(join(17983..18085,18189..18257,18683..18794))
                   /gene="bA576C12.1"
                   /pseudo
CDS               complement(join(17983..18085,18189..18257,18683..18794))
                   /gene="bA576C12.1"
                   /note="bA576C12.1 (VPS33B pseudogene)"
                   match: proteins: Tr:Q9NSF3 Tr:Q63616"
                   /pseudo
                   /codon_start=1
                   /evidence=not_experimental
repeat_region     18376..18594
                   /note="L1MA2 repeat: matches 6088..6308 of consensus"
repeat_region     19078..19134
                   /note="L2 repeat: matches 2647..2706 of consensus"
repeat_region     19152..19244
                   /note="MER5A repeat: matches 87..189 of consensus"
repeat_region     19623..19833
                   /note="L1MC/D repeat: matches 5412..5801 of consensus"
repeat_region     20152..20280
                   /note="AluJo/FRAM repeat: matches 165..293 of consensus"
repeat_region     20325..20542
                   /note="MIR repeat: matches 7..238 of consensus"
repeat_region     22120..22248
                   /note="L2 repeat: matches 2619..2750 of consensus"
repeat_region     23364..23653
                   /note="AluSx repeat: matches 1..292 of consensus"
repeat_region     24037..24135

```

```

repeat_region     24538..24614
                   /note="MIR repeat: matches 168..247 of consensus"
repeat_region     25175..25283
                   /note="FLAM_C repeat: matches 1..109 of consensus"
repeat_region     25519..25798
                   /note="AluSp repeat: matches 1..295 of consensus"
repeat_region     26291..26428
                   /note="MER5A repeat: matches 15..184 of consensus"
repeat_region     26552..26854
                   /note="AluSq repeat: matches 1..304 of consensus"
repeat_region     27292..27587
                   /note="AluY repeat: matches 1..295 of consensus"
repeat_region     27815..28006
                   /note="AluY repeat: matches 1..192 of consensus"
repeat_region     28021..28470
                   /note="L1R10A repeat: matches 13..548 of consensus"
repeat_region     28478..28582
                   /note="AluY repeat: matches 188..292 of consensus"
repeat_region     28585..28812
                   /note="MIR repeat: matches 8..256 of consensus"
repeat_region     29785..29914
                   /note="FLAM_C repeat: matches 1..132 of consensus"
repeat_region     30054..30368
                   /note="AluJo repeat: matches 1..312 of consensus"
repeat_region     30465..31203
                   /note="LIMB7 repeat: matches 5377..6168 of consensus"
repeat_region     31583..31900
                   /note="AluSx repeat: matches 1..301 of consensus"
misc_feature      complement(31945..31963)
                   /note="Sequence from overlapping clone bA276E15
                   (A1390067). Assembly confirmed by restriction digest."
repeat_region     31995..32288
                   /note="AluSp repeat: matches 1..295 of consensus"
repeat_region     33495..33581
                   /note="MIR repeat: matches 9..205 of consensus"
misc_feature      complement(33899..34718)
                   /note="match: GSS: Em:AQ743239"
repeat_region     33999..34064
                   /note="33 copies 2 mer tt 66% conserved"
repeat_region     34080..34720
                   /note="L2 repeat: matches 2080..2710 of consensus"
repeat_region     34993..35270
                   /note="AluJo repeat: matches 1..274 of consensus"
misc_feature      complement(36692..37430)
                   /note="match: GSS: Em:AQ481167"
repeat_region     37056..37286
                   /note="MIR repeat: matches 1..232 of consensus"
repeat_region     37287..37427
                   /note="L2 repeat: matches 1267..1415 of consensus"

Query Match      2.8%; Score 71; DB 9; Length 76114;
Best Local Similarity 100.0%; Pred.No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2050 TCGAATCTCTGATCGATTCACCCACCTTGCTCCCAAGTGTGGATTACAG 2109
Db 31664 TCGAATCTCTGATCGATTCACCCACCTTGCTCCCAAGTGTGGATTACAG 31605

Oy 2110 GTGTGAGCCAC 2120
Db 31604 GTGTGAGCCAC 31594

RESULT 20
AL354986/c 89557 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone R5-1100D14 map p34.1-35.***
DEFINITION SEQUENCING IN PROGRESS ***, 8 unordered pieces.
ACCESSION AL354986
VERSION AL354986.4 GI:11229132
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.

```

## ORGANISM

Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 88557)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

McLay, K.  
Direct Submission  
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Nov 20, 2000 this sequence version replaced gi:9797233.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: d1100D14

## COMMENT

----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator ABI; 2% of reads  
Chemistry: Dye-terminator Big Dye; 84% of reads  
Chemistry: Dye-terminator ET-amersham; 13% of reads  
Quality: 8653 bases at least Q40  
Consensus quality: 87145 bases at least Q30  
Insert size: 87857; sum-of-contigs  
Insert size: 109672; 31.9% error; agarose-fp  
Quality coverage: 5.81x in Q20 bases; sum-of-contigs Quality  
coverage: 4.81x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 8490: contig of 8490 bp in length  
\* 8491 8590: gap of 100 bp  
\* 8591 40289: contig of 31709 bp in length  
\* 40300 40399: gap of 100 bp  
\* 40400 45544: contig of 5145 bp in length  
\* 45545 45644: gap of 100 bp  
\* 45645 49137: contig of 3493 bp in length  
\* 49138 49237: gap of 100 bp  
\* 49238 53403: contig of 4166 bp in length  
\* 53404 53503: gap of 100 bp  
\* 53504 78346: contig of 24643 bp in length  
\* 78347 78446: gap of 100 bp  
\* 78447 86288: contig of 7842 bp in length  
\* 86289 86388: gap of 100 bp  
\* 86389 88557: contig of 2169 bp in length.  
Location/Qualifiers  
1. 88557

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p34.1-35"  
/clone="RP5-1100D14"  
/clone\_id="RPC1-5"  
1. 8490  
/note="assembly\_fragment:00310  
fragment\_chain:1"  
misc\_feature  
8591..40289  
/note="assembly\_fragment:00014  
fragment\_chain:1"  
misc\_feature  
40400..45544  
/note="assembly\_fragment:00898  
fragment\_chain:2"  
misc\_feature  
45645..49137  
/note="assembly\_fragment:00104

misc\_feature  
49238..53403  
/note="assembly\_fragment:00712"  
misc\_feature  
53504..78346  
/note="assembly\_fragment:01219"  
misc\_feature  
78447..86288  
/note="assembly\_fragment:00674  
fragment\_chain:3"  
misc\_feature  
86389..88557  
/note="assembly\_fragment:01199  
fragment\_chain:3  
clone end:17  
vector side:right"

misc\_feature

BASE COUNT 22574 a 21575 c 20359 g 23347 t 702 others  
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 88557;  
Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTCGACCTGACGATCCACCCACCTGAGCTCCCAAGGCTGGATTACAG 2109  
Db 68855 TCGAAGCTCGACCTGACGATCCACCCACCTGAGCTCCCAAGGCTGGATTACAG 68796

Qy 2110 GTGTGACCCAC 2120  
Db 68795 GTGTGACCCAC 68785

## RESULT 21

AC090114 91722 bp DNA linear PRI 13-MAR-2002  
LOCUS Homo sapiens BAC clone RP11-212P7 from 7, complete sequence.  
DEFINITION AC090114  
AC090114.5 GI:19387791  
VERSION HTG.  
KEYWORDS  
KEYWORDS  
SOURCE  
Homo sapiens  
ORGANISM

REFERENCE  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 91722)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
JOURNAL MEDLINE  
99063792  
PUBMED 9847074

REFERENCE  
AUTHORS 2 (bases 1 to 91722)  
Majumder, J., Meyer, R. and Doeber, A.  
The sequence of Homo sapiens BAC clone RP11-212P7  
JOURNAL Unpublished (2001)  
3 (bases 1 to 91722)  
Waterston, R.H.  
REFERENCE  
AUTHORS Direct Submission  
Submitted (16-FEB-2001) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 91722)  
Waterston, R.  
Direct Submission  
Submitted (13-MAR-2002) Department of Genetics, Washington  
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 13, 2002 this sequence version replaced gi:18151035.

COMMENT  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@wustl.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0212P07

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanehe, U. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-155G14. 2000 bp overlap. The clone sequenced to the right is RP13-488A5. Actual end of this clone is at base position 91722 of RP11-212P7.

#### FEATURES

```

source
    1..91722
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="7"
        /map="7"
        /clone="RP11-212P7"
        /clone_id="RPCT-11"
        /rpt_family="Alu"
        /rpt_family="Mariner"
repeat_region
    167..216
        /rpt_family="Alu"
repeat_region
    236..347
        /rpt_family="Alu"
misc_feature
    453..606
        /note="match to EST AL553530 (NID:g12893451)"
misc_feature
    453..602
        /note="match to EST AW957421 (NID:g8147224)"
misc_feature
    453..602
        /note="match to EST BF033256 (NID:g10740968)"
misc_feature
    453..602
        /note="similar to Mus musculus EST BG69684 (NID:g12552253)"
misc_feature
    457..602
        /note="similar to Homo sapiens EST AL520867 (NID:g12784360)"
misc_feature
    460..602
        /note="similar to EST BM452818 (NID:g18501858)"
misc_feature
    475..602
        /note="match to EST BG181401 (NID:g13703088)"
misc_feature
    502..602
        /note="match to EST BF332291 (NID:g11303039)"
misc_feature
    526..602
        /note="match to Bos taurus EST BM107596 (NID:g17038666)"
misc_feature
    530..602
        /note="match to EST AU118175 (NID:g10933201)"
misc_feature
    550..588
        /note="similar to Homo sapiens EST A1638878 (NID:g4691112)

```

```

repeat_region
    660..950
        /rpt_family="Alu"
repeat_region
    1286..1577
        /rpt_family="Alu"
repeat_region
    1583..1879
        /rpt_family="Alu"
misc_feature
    1803..2089
        /note="match to EST BF832110 (NID:g12180498)"
misc_feature
    1949..1952
        /note="match to EST N33555 (NID:g1153954) YY04F07.S1"
misc_feature
    2102
        /note="match to EST BF832110 (NID:g12180498)"
repeat_region
    2103..2411
        /rpt_family="Alu"
misc_feature
    2341..2699
        /note="similar to Homo sapiens EST AW812325 (NID:g7905319)"
misc_feature
    2435..2611
        /note="similar to Homo sapiens EST BF328761 (NID:g1299509)"
misc_feature
    2588..2592
        /note="match to EST AW352175 (NID:g6849888)"
misc_feature
    2646..2647
        /note="similar to Homo sapiens EST BF328761 (NID:g1299509)"
repeat_region
    2922..3215
        /rpt_family="Alu"
repeat_region
    3275..3295
        /rpt_family="AT-rich"
repeat_region
    3299..3394
        /rpt_family="Alu"
repeat_region
    3416..3454
        /rpt_family="TTTG)n"
repeat_region
    3457..3697
        /rpt_family="Alu"
misc_feature
    3725..4207
        /note="match to EST N33555 (NID:g1153954) YY04F07.S1"
misc_feature
    3943..3946
        /note="similar to Homo sapiens EST BG403117 (NID:g13296565)"
repeat_region
    3956..4016
        /rpt_family="L2"
repeat_region
    4134..4441
        /rpt_family="Alu"
misc_feature
    4443..4763
        /note="match to EST BE068262 (NID:g8412912)"
misc_feature
    4678..4801
        /note="similar to Homo sapiens EST A1638878 (NID:g4691112) t08d05.x1"
misc_feature
    4680
        /note="match to EST H98925 (NID:g1123593) YX31A12.S1"
misc_feature
    4684..4801
        /note="match to EST BE792117 (NID:g10213315)"
misc_feature
    4693..5154
        /note="match to EST H98925 (NID:g1123593) YX31A12.S1"
misc_feature
    4695..4840
        /note="match to EST AW957421 (NID:g8147224)"
misc_feature
    4695..4803
        /note="match to EST BF332291 (NID:g11303039)"
misc_feature
    4695..4801
        /note="match to EST AU118175 (NID:g10933201)"
misc_feature
    4695..4801
        /note="match to EST BF033256 (NID:g10740968)"
misc_feature
    4695..4801
        /note="match to EST BG181401 (NID:g13703088)"
misc_feature
    4695..4801
        /note="similar to Bos taurus EST BM107596 (NID:g17038666)"
misc_feature
    4695..4801
        /note="similar to EST BM452818 (NID:g18501858)"
misc_feature
    4695..4801
        /note="similar to Homo sapiens EST AL520867 (NID:g12784360)"

```

```

misc_feature      4695..4801
                  /note="similar to Homo sapiens EST BG403117
                  (NID:913296565)"
misc_feature      4695..4801
                  /note="similar to Mus musculus EST BG069684
                  (NID:912552253)"
misc_feature      4703..4801
                  /note="match to EST AA932787 (NID:93086752) coo6e04.s1"
misc_feature      4705..4746
                  /note="match to EST AL553530 (NID:912893451)"
misc_feature      4744..4801
                  /note="match to EST A1186139 (NID:93736777) qe3h10.s1"

Query Match      2.8%; Score 71; DB 9; Length 91722;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTCTGACCTGAGTGTATCCACCCACCTTGCTCCCAAGCTCGGATTACAG 2109
Db 58475 TCGAAGCTCTGACCTGAGTGTATCCACCCACCTTGCTCCCAAGCTCGGATTACAG 58534

Qy 2110 GTGTAGGCCAC 2120
Db 58535 GTGTAGGCCAC 58545

RESULT 22
AC023906/c 93287 bp DNA linear PRI 06-JUL-2001
LOCUS Homo sapiens chromosome 15 clone CTD-2184D3 map 15q21.2, complete
DEFINITION sequence.
ACCESSION AC023906
VERSION AC023906.7 GI:14595770
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 93287)
AUTHORS Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrision, G.,
James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T.,
Pate, D., and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 93287)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrision, G.,
James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T.,
Pate, D., and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 93287)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Neebitt, R.,
Pate, D., and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
REFERENCE 4 (bases 1 to 93287)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Neebitt, R.,
Pate, D., and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT ----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMMS
Web site: http://chroma.mbc.washington.edu/meg_www

```

```

Contact: leecrow@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

Note: Data from overlapping BACs AC010674 [drafting center:
UMMS], AC090970 [drafting center: UMMS], and AC016824 [drafting
center: GTC] was added for finishing
-----
FEATURES
source
Location/Qualifiers
1..93287
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.2"
/clone="CTD-2184D3"
/clone_lib="Cal Tech Human BAC library D"
/note="Data from overlapping clones CTD-2650P22 AC090970,
RP11-47K1 AC016824, and RP11-430B1 was added and the
consensus was determined from CTD-2184D3 to the extent
possible"
1..8893
/note="Overlap with CTD-2650P22 AC090970"
1..8893
/note="Overlap with RP11-47K1 AC016824"
6078..6140
/note="low quality data"
10529..10532
/note="low quality data"
14722..14806
/note="low quality data"
37627
/note="low quality data"
71552..71593
/note="low quality data"
82012..93287
/note="Overlap with RP11-430B1 AC010674"

BASE COUNT 25809 a 21125 c 20119 g 26234 t
ORIGIN

Query Match      2.8%; Score 71; DB 9; Length 93287;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTCTGACCTGAGTGTATCCACCCACCTTGCTCCCAAGCTCGGATTACAG 2109
Db 57434 TCGAAGCTCTGACCTGAGTGTATCCACCCACCTTGCTCCCAAGCTCGGATTACAG 57375

Qy 2110 GTGTAGGCCAC 2120
Db 57374 GTGTAGGCCAC 57364

RESULT 23
AP000098/c 100000 bp DNA linear PRI 25-MAY-2002
LOCUS Homo sapiens genomic DNA of 21q22.1, GART and AML related.
DEFINITION Q78C10-149C3 region, segment 1/20.
ACCESSION AP000098
VERSION AP000098.1 GI:4730832
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Shiba, T., and Sakaki, Y.
TITLE Homo sapiens 2,051,516bp genomic DNA of 21q22.1 GART and AML region
JOURNAL Published Only in Databases (1999)
REFERENCE 2 (bases 1 to 100000)
AUTHORS Hirakawa, M., Yamaguchi, H., Imai, K. and Shimada, Y.

```









```

repeat_region      /rpt_family="MIR"
                    complement(17712..17885)
repeat_region      /rpt_family="MERSA"
                    complement(18533..18714)
repeat_region      /rpt_family="MERS20"
                    complement(21055..21124)
repeat_region      /rpt_family="L2"
                    complement(21201..21250)
repeat_region      /rpt_family="L2"
                    complement(21659..21707)
repeat_region      /rpt_family="MIR"
                    complement(22379..22649)
repeat_region      /rpt_family="LMC1"
                    complement(22979..23292)
repeat_region      /rpt_family="AluSg"
                    complement(24216..24527)
repeat_region      /rpt_family="AluIo"
                    complement(24687..24980)
repeat_region      /rpt_family="L1PA4"
                    complement(25755..25782)
repeat_region      /rpt_family="AT_rich"
                    complement(25785..26070)
repeat_region      /rpt_family="AluSx"
                    complement(26071..26107)
repeat_region      /rpt_family="AT_rich"
                    complement(26186..26477)
repeat_region      /rpt_family="AluIo"
                    complement(26594..26757)
repeat_region      /rpt_family="MIR"
                    complement(26902..27202)
repeat_region      /rpt_family="AluSg"
                    complement(28890..29296)
repeat_region      /rpt_family="AluIo"
                    complement(29742..29896)
repeat_region      /rpt_family="MIR"
                    complement(30043..30592)
repeat_region      /rpt_family="MLTIF"
                    complement(31090..31367)
repeat_region      /rpt_family="AluSx"
                    complement(32074..32199)
repeat_region      /rpt_family="MIR"
                    complement(32206..32521)
repeat_region      /rpt_family="AluSg"
                    complement(32572..32757)
repeat_region      /rpt_family="MIR"
                    complement(32761..32838)
repeat_region      /rpt_family="MERS2"
                    complement(33407..33553)
repeat_region      /rpt_family="MERSB"
                    complement(33575..33756)
repeat_region      /rpt_family="MERSA"
                    complement(33771..33973)
repeat_region      /rpt_family="MERS8A"
                    complement(34187..34271)
repeat_region      /rpt_family="MIR"
                    complement(34988..35288)
repeat_region      /rpt_family="AluSx"
                    complement(35412..35477)
repeat_region      /rpt_family="MERSA"
                    complement(35620..36292)
repeat_region      /rpt_family="L1PA2"
                    complement(36302..36575)
repeat_region      /rpt_family="L1MB5"
                    complement(36577..36686)
repeat_region      /rpt_family="MIR"
                    complement(36933..37007)
repeat_region      /rpt_family="TGAAn"
                    complement(37875..37999)
repeat_region      /rpt_family="FLAM_A"
                    complement(38476..38556)
repeat_region      /rpt_family="PTR5"
                    complement(38557..38849)
repeat_region      /rpt_family="AluY"

```

```

repeat_region      38850..39247
                    /rpt_family="LRR12"
repeat_region      complement(39625..39748)
                    /rpt_family="MIR"
repeat_region      40650..40674
                    /rpt_family="CAG)n"
repeat_region      42135..42188
                    /rpt_family="(GGA)n"
repeat_region      complement(43811..43983)
                    /rpt_family="MIR"
repeat_region      44313..44493
                    /rpt_family="MERSB"
repeat_region      complement(44546..44620)
                    /rpt_family="MERS8A"
repeat_region      complement(45397..45480)
                    /rpt_family="L1MB8"
repeat_region      45517..45584
                    /rpt_family="MLTIF"
repeat_region      complement(45696..45804)

```

Query Match 2.8% Score 71; DB 9; Length 104147;  
 Best Local Similarity 100.0%; Pred. No. 1,2e-28; Mismatches 0; Gaps 0;

Qy 2050 TCGAATCTGACCTGAGTATACCCACCTTGCCCTCCCAAGTGTGGATTACAG 2109  
 Db 90800 TCGAATCTGACCTGAGTATACCCACCTTGCCCTCCCAAGTGTGGATTACAG 90741

Qy 2110 GTGTAGCCAC 2120  
 Db 90740 GTGTAGCCAC 90730

RESULT 27  
 LOCUS AC091614 107764 bp DNA linear PRI 22-DEC-2001  
 DEFINITION Homo sapiens chromosome 1 clone RP4-665J23, complete sequence.  
 AC091614 AL356854  
 VERSION AC091614.3 GI:1977203  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
 Haugen, E.D.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 107764)  
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and  
 Haugen, E.D.

TITLE Direct Submission  
 JOURNAL Submitted (09-MAY-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 REFERENCE 3 (bases 1 to 107764)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
 Haugen, E.D.

TITLE Direct Submission  
 JOURNAL Submitted (22-DEC-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 COMMENT On Dec 22, 2001 this sequence version replaced gi:15145523.

----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: http://www.genome.washington.edu  
 Contact: uwgchgs@u.washington.edu  
 Drafting Center: SC

----- Project Information  
 Center project name: chr-1  
 Center clone name: RP4-665J23 (sc02023)  
 ----- Summary Statistics  
 Sequencing vector: plasmid; U08752; 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap, version 0.990319  
 Consensus quality: 107697 bases at least Q40  
 Consensus quality: 107756 bases at least Q30  
 Consensus quality: 107764 bases at least Q20  
 Insert size: 107581; sum-of-coverage  
 Quality coverage: 8.7x in Q20 bases; sum-of-coverage

## Overlapping Sequences:

5' : Mapping in progress  
 3' : RP5-924G13 AL136109

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

## ECORI

## HindIII

## BglII

SeqDerMap	FingerPrint	SeqDerMap	FingerPrint	SeqDerMap	FingerPrint
-----	-----	-----	-----	-----	-----
6141	6207	2186	2167	9719	9643
-----	-----	-----	-----	-----	-----
2184	2186	449	<800	5671	5706
-----	-----	-----	-----	-----	-----
8065	8225	512	<800	4252	4269
-----	-----	-----	-----	-----	-----
486	<800	2814	2829	4573	4548
-----	-----	-----	-----	-----	-----
106	<800	1247	1203	1388	1389
-----	-----	-----	-----	-----	-----
5839	5881	4695	4744	8582	8585
-----	-----	-----	-----	-----	-----
5024	5124	3373	3404	2326	2507
-----	-----	-----	-----	-----	-----
843	843	3940	3959	12131	12076
-----	-----	-----	-----	-----	-----
15970	15803	976	964	1392	1389
-----	-----	-----	-----	-----	-----
2767	2784	513	<800	1410	1389
-----	-----	-----	-----	-----	-----
676	<800	612	<800	7569	7617
-----	-----	-----	-----	-----	-----
19591	19553	5767	5877	6505	6523
-----	-----	-----	-----	-----	-----
23557	23906	2059	2044	2944	2954
-----	-----	-----	-----	-----	-----

1568	1551	850	862	6283	6264
-----	-----	-----	-----	-----	-----
5999	5881	10247	10174	580	<800
-----	-----	-----	-----	-----	-----
5142	5124	4011	3959	743	748
-----	-----	-----	-----	-----	-----
121	<800	3566	3557	23	<800
-----	-----	-----	-----	-----	-----
18437	18067	872	862	14146	14238
-----	-----	-----	-----	-----	-----
1250	1234	4776	4744	2394	2507
-----	-----	-----	-----	-----	-----
5883	5877	1672	1616	3125	3125
-----	-----	-----	-----	-----	-----
10636	10530	3196	3196	2507	2507
-----	-----	-----	-----	-----	-----
8601	8587	236	<800	2507	2507
-----	-----	-----	-----	-----	-----
2088	2167	2507	2507	2507	2507
-----	-----	-----	-----	-----	-----
2110	2116	1312	1295	5472	5472
-----	-----	-----	-----	-----	-----
7406	7332	5546	5472	2507	2507
-----	-----	-----	-----	-----	-----
241	<800	2508	1034	4814	4814
-----	-----	-----	-----	-----	-----
3917	3959	1048	4814	4269	4269
-----	-----	-----	-----	-----	-----
859	862	4851	4269	3583	3583
-----	-----	-----	-----	-----	-----
166	<800	4330	335	<800	<800
-----	-----	-----	-----	-----	-----
3133	3108	3594	335	373	373
-----	-----	-----	-----	-----	-----
2682	2707	335	373	373	373
-----	-----	-----	-----	-----	-----
392	<800	2829	2829	373	373
-----	-----	-----	-----	-----	-----
2814	2829	373	373	373	373
-----	-----	-----	-----	-----	-----
373	<800	373	373	373	373

## FEATURES

## source

Location/Qualifiers  
 1..107764  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP4-65J23"  
 /clone\_1fb="RP4-65J23"  
 /note="single subclone region"

## misc\_feature

45711..45818  
 /note="single subclone region"

## BASE COUNT

33346 a 21787 c 20575 g 32056 t

Query Match 2.8%; Score 71; DB 9; Length 107764;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## ORIGIN

2050 TCGAATCTGACCTGATGATCCACCCACCTTGAGCTCCCAAGTGTGGATTACAG 2109

DB 49247 TCGAATCTGACCTGATGATCCACCCACCTTGAGCTCCCAAGTGTGGATTACAG 49306

QY 2110 GTGTGAGCCAC 2120

DB 49307 GTGTGAGCCAC 49317

## RESULT 28

AC131236/c 108239 bp DNA linear HTG 19-AUG-2002

## LOCUS

AC131236 Homo sapiens clone CTD-2260B7, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 14

## DEFINITION

unordered pieces.

## ACCESSION

AC131236

VERSION  
AC131236.1 GI:22297331  
KEYWORDS  
HTG, HTGS\_PHASE1.  
SOURCE  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C., Albrook, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T., Barbieri, J., Benson, J., Bimaga, K., Blankenburg, K., Bonini, D., Bouck, J., Bowe, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Doukwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.U., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M., Faller, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B., Homei, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryak, J., Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loussag, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Maeser, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokemko, S., Ogun, M., Okwodu, G., Oreguine, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, J., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, K., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Uman, K., Vaequez, U., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Woodcock, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT  
Unpublished  
2 (bases 1 to 108239)  
Direct Submission  
Submitted (19-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project Information  
Center project name: HELI  
Center clone name: CTD-226087  
Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap, version 0.99029  
Consensus quality: 9784 bases at least Q40  
Consensus quality: 101040 bases at least Q30  
Consensus quality: 103062 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2543: contig of 2543 bp in length  
2544 2643: gap of unknown length  
2644 5633: contig of 2990 bp in length  
5634 5733: gap of unknown length  
5734 9343: contig of 3610 bp in length  
9344 9443: gap of unknown length  
9444 12822: contig of 3378 bp in length  
12822 12921: gap of unknown length  
12921 18419: contig of 5498 bp in length  
18420 18519: gap of unknown length  
18520 25079: contig of 6560 bp in length  
25080 25179: gap of unknown length  
25180 35297: contig of 10118 bp in length  
35298 35397: gap of unknown length  
35398 42548: contig of 7151 bp in length  
42549 42648: gap of unknown length  
42649 52259: contig of 9611 bp in length  
52260 52359: gap of unknown length  
52360 65801: contig of 13442 bp in length  
65802 65901: gap of unknown length  
65902 72723: contig of 6822 bp in length  
72724 72824: gap of unknown length  
72825 82707: contig of 9884 bp in length  
82708 82807: gap of unknown length  
82808 85092: contig of 12285 bp in length  
85093 95192: gap of unknown length  
95193 108239: contig of 13047 bp in length.  
Location/Qualifiers  
1..108239  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CMD-226087"

BASE COUNT 28670 a 26415 c 24912 g 26698 t 1544 others  
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 108239;  
Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCGACCTGAGTATCCACCCAGCTGCTCCCAAGCTGGATTACAG 2109  
|||||  
DB 3860 TCGAAGCTCGACCTGAGTATCCACCCAGCTGCTCCCAAGCTGGATTACAG 3801  
|||||  
QY 2110 GTGTAGCCAC 2120  
|||||  
DB 3800 GTGTAGCCAC 3790  
|||||

RESULT 29  
AL390208/c 109878 bp DNA linear PRI 20-JAN-2001  
LOCUS  
DEFINITION Human DNA sequence from clone Rpl1-787122 on chromosome 6, complete  
sequence.  
ACCESSION AL390208  
VERSION AL390208.17 GI:12331104  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 109878)  
AUTHORS Blakey, S.  
TITLE Direct Submission  
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humgeny@sanger.ac.uk](mailto:humgeny@sanger.ac.uk) Clone

## COMMENT

requests: clonerequest@sanger.ac.uk  
On Jan 22, 2001 this sequence version replaced gi:12329420.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

RP11-787122 is from the library RPCT-11.3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-787122. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-787122 is at 1 in this sequence. The true right end of clone RP11-487P23 is at 109779 in this sequence. The true right end of clone RP3-397N18 is at 97339 in this sequence.

## FEATURES

Location/Qualifiers

1..109878  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-787122"  
/clone\_1fb="RPCT-11.3"  
misc\_feature 83165..83200

BASE COUNT 31786 a 21771 c 21950 g 34371 t  
ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 109878;  
Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCCCAAGTGTGGATTACAG 2109

DB 44625 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCCCAAGTGTGGATTACAG 44566

QY 2110 GTGTGAGCCAC 2120

DB 44565 GTGTGAGCCAC 44555

RESULT 30  
AC107310.2/c  
WPCOMMENT

Sequence split into 4 fragments LOCUS AC107310 Accession AC107310

Fragment Name	Begin	End
AC107310.0	1	110000
AC107310.1	100001	210000
AC107310.2	200001	310000
AC107310.3	300001	357311

Continuation (3 of 4) of AC107310 from base 200001 (AC107310 Homo sapiens chromosome 3

Query Match 2.8%; Score 71; DB 2; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCCCAAGTGTGGATTACAG 2109  
DB 44603 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCCCAAGTGTGGATTACAG 44544  
QY 2110 GTGTGAGCCAC 2120  
DB 44543 GTGTGAGCCAC 44533

## RESULT 31

AL357500

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

On Nov 16, 2001 this sequence version replaced gi:16944735. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

RP11-242024 is from the library RPCT-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-242024. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP11-242024 is at 112392 in this sequence. The true right end of clone RP4-604A21 is at 2000 in this sequence.

## FEATURES

Location/Qualifiers

1..112392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-242024"  
/clone\_1fb="RPCT-11.1"  
17380..17440

misc\_feature

/note="Sequence from overlapping clone RP11-467D18(AL590729). Assembly confirmed by restriction digest."

misc\_feature 57015..57148  
/note="Sequence from overlapping clone





Mapping Group. Further information can be found at <http://www.sanger.ac.uk/RSB/Chr20>  
IMPORTANT: This sequence is not the entire insert of clone RPS-927M24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPS-927M24 is at 1 in this sequence. The true left end of clone RPS-1100H13 is at 121823 in this sequence. The true right end of clone RPS-2189E is at 65561 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPS-927M24 is from the library RPS-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pCYFAC2.

## FEATURES

source  
1. .121922  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="RPS-927M24"  
/clone\_11b="RPS-5"  
1. .248  
/note="AluX repeat: matches 59. .306 of consensus"  
283. .318  
/note="AluU repeat: matches 1. .36 of consensus"  
402. .659  
/note="AluU repeat: matches 43. .297 of consensus"  
1173. .1465  
/note="AluX repeat: matches 1. .299 of consensus"  
/complement(1611. .2023)  
/note="match: GSS: Em:AQ428420"  
1808. .1884  
/note="MER91C repeat: matches 63. .140 of consensus"  
2584. .2777  
/note="MT1A1 repeat: matches 179. .356 of consensus"  
2898. .3524  
/note="CPG island"  
/evidence="not\_experimental"  
3023. .7234  
/gene="dJ927M24.1"  
/gene="dJ927M24.1"  
join(<3023. .3325,3779. .3936,4975. .5043,6649. .6760,6846. .57234)  
/gene="dJ927M24.1"  
/product="dJ927M24.1.3 (Novel) protein, isoform 3)"  
/note="match: ESTs: Em:BE747933 Em:BE663360 Em:BE883052 Em:BE277445 Em:BE277383 Em:BE730060 Em:BE395751"  
/evidence="not\_experimental"  
join(<3037. .3325,3779. .3936,4243. .4406,4975. .5043,6649. .6760,6846. .57234)  
/gene="dJ927M24.1"  
/product="dJ927M24.1.2 (Novel) protein, isoform 2)"  
/note="match: ESTs: Em:AM753563 Em:R13556 Em:R14470 Em:R87094 Em:AA827892 Em:T91926 Em:H40444  
match: proteins: Sw:P22398"  
/evidence="not\_experimental"  
3564. .3690  
/note="MIR repeat: matches 108. .238 of consensus"  
join(<3637. .3936,4975. .5043,6649. .6735)  
/gene="dJ927M24.1"  
/product="dJ927M24.1.1 (Novel) protein, isoform 1)"  
/note="match: ESTs: Em:H26725 Em:H40444 Em:R46422 Em:H22204 Em:A1808326 Em:A115783 Em:R87094 Em:A464847 Em:A181822 Em:A1003178 Em:AW515615 Em:T91926 Em:A1924555 Em:AW293937 Em:A1276171 Em:AA777670 Em:AW510700 Em:A1084790 Em:AA411324 Em:AW189888 Em:A1760566 Em:A1275468 Em:R13596 Em:T25024 Em:AW628424 Em:AA468277 Em:R14470 Em:AW375363 Em:AA468204 Em:AA470331 Em:AA491735 Em:AA468450 Em:AA470347 Em:AA827892"

misc\_feature  
5628. .6153  
/gene="dJ927M24.1"  
/note="match: STS: Em:HS83JL11T"  
6087. .6222  
/note="2 copies 68 mer 98% conserved"  
/complement(6846. .7233)  
/note="match: STS: Em:G22570"  
7216. .7221  
/gene="dJ927M24.1"  
7234  
polyA\_site  
7965. .8266  
/note="AluX repeat: matches 1. .306 of consensus"  
9033. .9104  
/note="MIR repeat: matches 176. .248 of consensus"  
9268. .9491  
/note="U2 repeat: matches 2295. .2519 of consensus"  
9658. .10153  
/note="AluY repeat: matches 1. .296 of consensus"  
10751. .11079  
/note="AluX repeat: matches 1. .311 of consensus"  
11082. .11386  
/note="AluU repeat: matches 1. .302 of consensus"  
11658. .11789  
/note="FLAN A repeat: matches 1. .132 of consensus"  
/complement(11915. .12413)  
/note="match: GSS: Em:AQ016485"  
12147. .12332  
/note="MIR repeat: matches 61. .254 of consensus"  
12862. .12935  
/note="U2 repeat: matches 2594. .2674 of consensus"  
12965. .13052  
/note="U2 repeat: matches 2657. .2750 of consensus"  
13070. .13250  
/note="MER33 repeat: matches 138. .324 of consensus"  
13251. .13533  
/note="AluX repeat: matches 1. .286 of consensus"  
13534. .13687  
/note="MER33 repeat: matches 1. .138 of consensus"  
13958. .14171  
/note="U2 repeat: matches 2483. .2720 of consensus"  
14269. .14384  
/note="MIR repeat: matches 23. .147 of consensus"  
14410. .14701  
/note="AluX repeat: matches 1. .291 of consensus"  
14732. .15092  
/note="L1MD3 repeat: matches 7374. .7739 of consensus"  
15093. .15267  
/note="AluX repeat: matches 136. .308 of consensus"  
15268. .15572  
/note="AluX repeat: matches 1. .307 of consensus"  
15573. .15707  
/note="AluX repeat: matches 1. .136 of consensus"  
15708. .15917  
/note="L1MD3 repeat: matches 7166. .7374 of consensus"  
16168. .16474  
/note="AluX repeat: matches 1. .307 of consensus"  
16520. .16642  
/note="L1M4 repeat: matches 2908. .3027 of consensus"  
16978. .17945  
/note="SVA repeat: matches 1. .954 of consensus"  
17532. .18747  
/note="CPG island"  
/evidence="not\_experimental"  
17946. .18046  
/note="SVA repeat: matches 630. .730 of consensus"  
18048. .18918  
/note="SVA repeat: matches 521. .1386 of consensus"  
18961. .19247  
/note="AluU repeat: matches 1. .309 of consensus"  
19273. .19302  
/note="15 copies 2 mer at 90% conserved"

```

repeat_region      19305..19404
                    /note="50 copies 2 mer ta 80% conserved"
repeat_region      19406..19481
                    /note="38 copies 2 mer ta 85% conserved"
repeat_region      19518..20741
                    /note="18 copies 68 mer 59% conserved"
repeat_region      20339..20774
                    /note="218 copies 2 mer ta 75% conserved"
repeat_region      20354..20777
                    /note="11 copies 38 mer 74% conserved"
repeat_region      20777..20908
                    /note="AluIb repeat: matches 1..131 of consensus"
repeat_region      20909..21213
                    /note="AluSg repeat: matches 1..304 of consensus"
repeat_region      21214..21374
                    /note="AluIb repeat: matches 131..295 of consensus"
repeat_region      21413..21722
                    /note="AluSx repeat: matches 3..312 of consensus"
misc_feature       22040..22501
                    /note="match: GSS: Em:A0594574"
repeat_region      22195..22256
                    /note="MER91 repeat: matches 1..62 of consensus"

```

```

Query Match      2.8%; Score 71; DB 9; Length 121922;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGAGTGATCCACCACTTGCGCTCCCAAGTCTGGATTACAG 2109
      |||
Db 47041 TCGAAGCTCTGACCTGAGTGATCCACCACTTGCGCTCCCAAGTCTGGATTACAG 47100
      |||
QY 2110 GTGTGAGCCAC 2120
      |||
Db 47101 GTGTGAGCCAC 47111

```

```

RESULT 34
AL162716/c      122916 bp      DNA      linear      PRI 18-JAN-2002
LOCUS           Human DNA sequence from clone Rpl1-339124 on chromosome 13.
DEFINITION      complete sequence.
ACCESSION       AL162716
VERSION         AL162716.9      GI:18375788
KEYWORDS        HTG.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 122916)
REFERENCE
AUTHORS        Tyomans, A.
TITLE          Direct Submision
JOURNAL        Submitted (18-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequests@sanger.ac.uk
On Jan 25, 2002 this sequence version replaced gi:14132865.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submision
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

```

was generated from part of bacterial clone contigs of human Chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>  
Rpl1-339124 is from the library RPl-11.2 constructed by the group of Pletier de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6  
IMPORTANT: This sequence is not the entire insert of clone Rpl1-339124. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true right end of clone Rpl1-339124 is at 122916 in this sequence. The true right end of clone Rpl1-16815 is at 100 in this sequence.

```

FEATURES
SOURCE          1..122916
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="13"
                /clone="Rpl1-339124"
                /clone_1b="RPl-11.2"
                /clone_1b="RPl-11.2"
BASE COUNT      38400 a 22253 c 21859 g 40404 t
ORIGIN

```

```

Query Match      2.8%; Score 71; DB 9; Length 122916;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGAGTGATCCACCACTTGCGCTCCCAAGTCTGGATTACAG 2109
      |||
Db 5002 TCGAAGCTCTGACCTGAGTGATCCACCACTTGCGCTCCCAAGTCTGGATTACAG 4943
      |||
QY 2110 GTGTGAGCCAC 2120
      |||
Db 4942 GTGTGAGCCAC 4932

```

```

RESULT 35
CNS01DMK/c      125780 bp      DNA      linear      PRI 23-MAY-2001
LOCUS           Human chromosome 14 DNA sequence BAC C-2246N19 of library CalTech-D
DEFINITION      Human chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION       AL137229
VERSION         AL137229.4      GI:14268350
KEYWORDS        HTG; HTGS_ACTIVEPIN.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 125780)
REFERENCE
AUTHORS        Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brothier, P., Catolico, L., Barbe, V., Pelleter, E., Artiguenave, F.,
Levy, M., Eckembey, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissenbach, J.
TITLE          Sequencing of the human chromosome 14
JOURNAL        Unpublished
REFERENCE
AUTHORS        Genoscope.
TITLE          Direct Submision
JOURNAL        Submitted (23-MAY-2001) Genoscope - Centre National de Sequenage;
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
On May 30, 2001 this sequence version replaced gi:11875955.
----- Genome Center
Center: Genoscope / Centre National de Sequenage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-3059L23
Downstream BAC (overlapping the SP6 end) : C-2555C10 (AC-AL136293)
----- Summary Statistics

```



Assembly program: Phrap; version 2.0  
Quality coverage: 7.96x in Q20 bases; sum-of-contigs

-----  
Overall quality chart :  
Range : bases

```

0 - 9 :
10 - 19 :
20 - 29 :
30 - 39 :
40 - 49 :
50 - 59 :
60 - 69 :
70 - 79 :
80 - 89 :
90 - 99 :

```

Percentage of bases with a quality value >= 40 : 99 %.

# FEATURES

source

Location/Qualifiers

1. 125780

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="14"

/clone\_lib="Caltech-D"

14526.14736

/note="matching EMBL:R49500"

RHdb:RH26306

dbSTS:STS7102

Identified using the e-PCR software (G. Schuler)

14747.14874

/note="matching EMBL:D59403"

RHdb:RH47332

dbSTS:STS40392

Identified using the e-PCR software (G. Schuler)

21680.21791

/note="matching EMBL:H63409"

RHdb:RH53932

dbSTS:STS27684

Identified using the e-PCR software (G. Schuler)

21701.21853

/note="matching EMBL:H63409"

RHdb:RH44578

dbSTS:STS37646

Identified using the e-PCR software (G. Schuler)

33434 a 28298 c 29699 g 34349 t

BASE COUNT

ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 125780;  
Best Local Similarity 100.0%; Fred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGAGCTGAGTATCCACCCAGCTGCTCCCAAGTCTGGATTACAG 2109  
DB 74618 TCGAAGCTCTGAGCTGAGTATCCACCCAGCTGCTCCCAAGTCTGGATTACAG 74559  
QY 2110 GTGAGACCCAC 2120  
DB 74558 GTGAGACCCAC 74548

RESULT 36  
AC004822/c 127824 bp DNA linear PRI 21-DEC-1999  
LOCUS AC004822 Homo sapiens PAC clone RPI-170D19 from Xg23, complete sequence.  
DEFINITION AC004822 Homo sapiens PAC clone RPI-170D19 from Xg23, complete sequence.  
ACCESSION AC004822  
VERSION AC004822.1 GI:3845420  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 127824)  
AUTHORS Kalicki, J. and Mead, K.  
TITLE The sequence of Homo sapiens PAC clone RPI-170D19  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 127824)  
TITLE Waterston, R. H.  
JOURNAL Direct Submission  
Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 3 (bases 1 to 127824)  
AUTHORS Waterston, R.  
JOURNAL Direct Submission  
Submitted (05-NOV-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
4 (bases 1 to 127824)  
AUTHORS Waterston, R.  
JOURNAL Direct Submission  
Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 5, 1998 this sequence version replaced GI:3309103.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
Center project name: H\_DU0170D19  
-----  
NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.  
-----  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.  
-----  
MAPPING INFORMATION:  
This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by David Bentley's chromosome X  
mapping group at the Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, UK. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>  
-----  
SOURCE INFORMATION:  
This clone was derived from human PAC library RPI-1, prepared by  
Pieter de Jong and coworkers at the Roswell Park Cancer Institute  
(<http://bacpac.med.buffalo.edu>) using the method described by  
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from  
one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(<http://www.genomesystems.com>) or Research Genetics, Inc.  
(<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2  
NEIGHBORING SEQUENCE INFORMATION:  
The actual end of clone RPI-170D19 is at base position 1 of  
RPI-170D19; the actual end is at 127824 of RPI-170D19.  
-----  
Location/Qualifiers  
1. 127824  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xg23"  
/clone="RPI-170D19"  
/clone\_lib="RPI-1"  
38.465  
/db\_xref="GI:2624100"  
991..2104  
repeat\_region

```

repeat_region      25709. .26893
                    /rpt_family="n1"
repeat_region      26911. .27044
                    /rpt_family="n1"
repeat_region      27097. .27126
                    /rpt_family="AT_rich"
repeat_region      27962. .27993
                    /rpt_family="CATAA)n"
repeat_region      28726. .28829
                    /rpt_family=" (TAAA)n"
repeat_region      30880. .30902
                    /rpt_family=" (CAA)n"
repeat_region      31045. .31155
                    /rpt_family="MIR"
repeat_region      31326. .31455
                    /rpt_family="Alu"
repeat_region      31456. .31516
                    /rpt_family="purine-rich"
repeat_region      31524. .31628
                    /rpt_family=" (GAA)n"
repeat_region      31847. .31922
                    /rpt_family="MIR"
repeat_region      32066. .32377
                    /rpt_family="Alu"
repeat_region      34619. .34735
                    /rpt_family=" (TA)n"
repeat_region      34736. .34759
                    /rpt_family=" (GA)n"
repeat_region      34847. .35109
                    /rpt_family="MIR"
repeat_region      35446. .35485
                    /rpt_family="AT_rich"
misc_feature       35676. .35854
                    /note="match to EST AA682775 (NTD:92669458) ah52h07.s1"
repeat_region      35792. .36098
                    /rpt_family="Alu"
repeat_region      36614. .36698
                    /rpt_family="L2"
repeat_region      38042. .38334
                    /rpt_family="Alu"
misc_feature       38235. .38521
                    /note="match to EST AA682775 (NTD:92669458) ah52h07.s1"
repeat_region      38622. .38655
                    /rpt_family=" (CA)n"
repeat_region      39148. .39201

```

Query Match Best Local Similarity 2.8%; Score 71; DB 9; Length 127824;  
 Query Match Best Local Similarity 100.0%; Pred. No. 1,2e+28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCTCGAGCTCAGGTGATTCACCCAGCTGGCTCCCAAGTCMGAGATTAAAG 2109  
 Db 120490 TCGAATCTCGAGCTCAGGTGATTCACCCAGCTGGCTCCCAAGTGTGGAATTACAG 120431  
 |||||  
 QY 2110 GTGTGAGCCAC 2120  
 Db 120430 GTGTGAGCCAC 120420  
 |||||

RESULT	37
AC069548/c	
LOCUS	AC069548
DEFINITION	Homo sapiens chromosome 10 clone RP11-522H2, complete sequence.
ACCESSION	AC069548
VERSION	AC069548.6 GI:20087107
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (Baes I to J19095)
AUTHORS	Smith,D.R.
TITLE	Genome Therapeutics Corporation Sequencing Center: Human Genome

## Sequence Data

JOURNAL  
REFERENCE 2 (bases 1 to 131909)

AUTHORS  
TITLE  
JOURNAL

Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA  
3 (bases 1 to 131909)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (26-FEB-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA  
4 (bases 1 to 131909)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA  
5 (bases 1 to 131909)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA  
On Apr 9, 2002 this sequence version replaced gi:19774381.

COMMENT

Center: Genome Therapeutics Corporation  
Center code: GTC  
Web site: <http://www.genomecorp.com/>  
Contact: [gtrc-sequence@genomecorp.com](mailto:gtrc-sequence@genomecorp.com)  
Project Information  
Center project name: hg231

IMPORTANT: This sequence is not the entire insert of clone RPII-522H2. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RPII-354E23 is at 129909 in this sequence.

----- Summary Statistics -----  
Sequencing vector: N/A  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 990315

## FEATURES

source  
Location/Qualifiers  
1. 131909  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RPII-522H2"  
/clone\_11b="RPCI-11"  
BASE COUNT 40940 a 26901 c 26639 g 37429 t  
ORIGIN

Query Match 2.8%; Score 71; DB 9; Length 131909;  
Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTGACCTGAGTGATCCACCACTGGCTGCCAAGTGGATTACAG 2109  
Db 85510 TCGAAGCTGACCTGAGTGATCCACCACTGGCTGCCAAGTGGATTACAG 85451

QY 2110 GTGTGAGCCAC 2120  
Db 85450 GTGTGAGCCAC 85440

RESULT 38  
AL589792 132592 bp DNA linear HTG 10-JUL-2001  
LOCUS  
DEFINITION Homo sapiens chromosome 1, clone RP5-889P23 map p36.21-36.33, \*\*\*  
ACCESSION AL589792  
VERSION AL589792.1 GI:13277194  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.

## SOURCE

human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 132592)

AUTHORS  
TITLE  
JOURNAL  
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone  
request: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

COMMENT

----- Genome Center -----  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
Project Information  
Center project name: d1889P23  
----- Summary Statistics -----  
Assembly program: XGAP4; version 4.5  
Sequencing vector: M13; M7815; 12% of reads  
Sequencing vector: plasmid; L08752; 87% of reads  
Chemistry: Dye-terminator ABI; 14% of reads  
Chemistry: Dye-terminator Big Dye; 60% of reads  
Chemistry: Dye-terminator Big Dye; 60% of reads  
Chemistry: Dye-terminator Big Dye; 60% of reads  
Chemistry: Dye-terminator Big Dye; 60% of reads  
Consensus quality: 131458 bases at least Q40  
Consensus quality: 131709 bases at least Q30  
Consensus quality: 131965 bases at least Q20  
Insert size: 132292; sum-of-contigs  
Insert size: 132129; 14.2% error; agarose-fp  
Quality coverage: 11.37x in Q20 bases; sum-of-contigs Quality  
coverage: 11.69x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3331: contig of 3331 bp in length.  
\* 3332 3431: gap of 100 bp  
\* 3432 90998: contig of 87567 bp in length  
\* 90999 91098: gap of 100 bp  
\* 91099 130106: contig of 39008 bp in length  
\* 130107 130206: gap of 100 bp  
\* 130207 132592: contig of 2386 bp in length.  
FEATURES  
source  
Location/Qualifiers  
1. 132592  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p36.21-36.33"  
/clone="RP5-889P23"  
/clone\_11b="RPCI-5"  
1. 3331  
/note="assembly\_fragment:00799"  
3432. 90998  
/note="assembly\_fragment:02114"  
91099. 130106  
/note="assembly\_fragment:03033.0"  
130207. 132592  
/note="assembly\_fragment:03380"  
BASE COUNT 26843 a 36833 c 36775 g 31841 t 300 others  
ORIGIN

misc\_feature  
/note="assembly\_fragment:00799"  
misc\_feature  
3432. 90998  
/note="assembly\_fragment:02114"  
91099. 130106  
/note="assembly\_fragment:03033.0"  
130207. 132592  
/note="assembly\_fragment:03380"  
BASE COUNT 26843 a 36833 c 36775 g 31841 t 300 others  
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 132592;  
Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2050 TCGAAGCTGACCTGAGTGATCCACCACTGGCTGCCAAGTGGATTACAG 2109  
|||||

Db 105811 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCTCCCAAGTGTGGATTACAG 105870  
 Oy 2110 GTGTGAGCCAC 2120  
 |||||  
 Db 105871 GTGTGAGCCAC 105881

RESULT 39  
 AC007644 141048 bp DNA linear HTG 23-MAY-1999  
 AC007644 Homo sapiens chromosome 17 clone hRPK.19 F.16 map 17, \*\*\*  
 DEFINITION SEQUENCING IN PROGRESS \*\*\*; 17 unordered pieces.  
 AC007644  
 AC007644.1 GI:4884012  
 HTG, HTGS\_PHASE1.  
 SOURCE  
 ORGANISM  
 Homo sapiens.  
 Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 141048)  
 2 (bases 1 to 141048)  
 Homo sapiens chromosome 17, clone hRPK.19\_F.16  
 Unpublished

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 AUTHORS  
 1 (bases 1 to 141048)  
 2 (bases 1 to 141048)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckwith, R., Berr, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collins, A., Cooke, P., Dearrellano, K., Depayre, E., Devon, K., Dewar, K., Donegan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardina, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karas, A., Lehotzky, J., Liu, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGuire, A., McKernan, K., McLaughlin, J., Melidim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Naylor, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Testa, S., Toruella-Miller, J., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
 Direct Submission  
 Submitted (23-MAY-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker: Smt, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 17 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 448: contig of 448 bp in length  
 449 2872: contig of 2424 bp in length  
 2873 6852: contig of 3980 bp in length  
 6853 10142: contig of 3290 bp in length  
 10143 14836: contig of 4694 bp in length  
 14837 22329: contig of 7493 bp in length  
 22330 27992: contig of 5663 bp in length  
 27993 33787: contig of 5795 bp in length  
 33788 40261: contig of 6474 bp in length  
 40262 45996: contig of 5735 bp in length  
 gap of unknown length

45997 53175: contig of 7179 bp in length  
 gap of unknown length  
 53176 59404: contig of 6229 bp in length  
 gap of unknown length  
 59405 69232: contig of 9828 bp in length  
 gap of unknown length  
 69233 78639: contig of 9407 bp in length  
 gap of unknown length  
 78640 94682: contig of 16043 bp in length  
 gap of unknown length  
 94683 127542: contig of 32860 bp in length  
 gap of unknown length  
 127543 141048: contig of 13506 bp in length.  
 Location/Qualifiers  
 1..141048  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="hRPK.19 F.16"  
 /clone.lib="RPCT-11 human BAC library"  
 BASE COUNT 38166 a 30217 c 31995 g 40108 t 562 others  
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 141048;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2050 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCTCCCAAGTGTGGATTACAG 2109  
 Db 134616 TCGAAGCTCTGAGCTGATGATCCACCCAGCTTGCTCCCAAGTGTGGATTACAG 134675

Oy 2110 GTGTGAGCCAC 2120  
 |||||  
 Db 134676 GTGTGAGCCAC 134686

RESULT 40  
 AC068266 149235 bp DNA linear HTG 24-AUG-2002  
 AC068266 Homo sapiens chromosome 2 clone RP11-146F22 map 2, WORKING DRAFT  
 DEFINITION SEQUENCE, 24 unordered pieces.  
 AC068266  
 AC068266.2 GI:8225539  
 HTG, HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 ORGANISM  
 Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 149235)  
 2 (bases 1 to 149235)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukington, B., Brown, A., Burkett, G., Campopiano, A., Caetle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J., LaRoque, K., Lamazeres, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R., Melidim, J., Menau, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,

## TITLE

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## JOURNAL

Submitted (30-Apr-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 149235)

## AUTHORS

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Zainoun, J., Zainoun, J., Zimmer, A. and Zody, M.

## TITLE

Submitted (24-Aug-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## JOURNAL

On Jun 3, 2000 this sequence version replaced gi:7671341.

## COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/XM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Genome Center

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

Project Information

Center project name: L9679

Center clone name: 146 F.22

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 138221 bases at least Q40

Consensus quality: 143619 bases at least Q30

Consensus quality: 145725 bases at least Q20

Insert size: 151000; agarose-fp

Insert size: 146935; sum-of-ctgigs

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-ctgigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2046: contig of 2046 bp in length  
\* 2047 2146: gap of 100 bp  
\* 2147 2522: contig of 376 bp in length  
\* 2523 2622: gap of 100 bp  
\* 2623 4106: contig of 1484 bp in length  
\* 4107 4206: gap of 100 bp  
\* 4207 6021: contig of 1815 bp in length  
\* 6022 6121: gap of 100 bp  
\* 6122 8689: contig of 2568 bp in length  
\* 8690 8789: gap of 100 bp  
\* 8790 11605: contig of 2816 bp in length

## FEATURES

## source

11606 11705: gap of 100 bp  
\* 11706 15022: contig of 3317 bp in length  
\* 15023 15122: gap of 100 bp  
\* 15123 18721: contig of 3599 bp in length  
\* 18722 18821: gap of 100 bp  
\* 18822 23727: contig of 4906 bp in length  
\* 23728 23827: gap of 100 bp  
\* 23828 29512: contig of 5685 bp in length  
\* 29513 29612: gap of 100 bp  
\* 29613 34992: contig of 5380 bp in length  
\* 34993 35092: gap of 100 bp  
\* 35093 38620: contig of 3528 bp in length  
\* 38621 38720: gap of 100 bp  
\* 38721 43918: contig of 5198 bp in length  
\* 43919 44018: gap of 100 bp  
\* 44019 49427: contig of 5409 bp in length  
\* 49428 49527: gap of 100 bp  
\* 49528 55671: contig of 6144 bp in length  
\* 55672 55771: gap of 100 bp  
\* 55772 61066: contig of 5295 bp in length  
\* 61067 61166: gap of 100 bp  
\* 61167 67660: contig of 6494 bp in length  
\* 67661 67760: gap of 100 bp  
\* 67761 73637: contig of 5877 bp in length  
\* 73638 73737: gap of 100 bp  
\* 73738 81886: contig of 8149 bp in length  
\* 81887 81986: gap of 100 bp  
\* 81987 92046: contig of 10060 bp in length  
\* 92047 92146: gap of 100 bp  
\* 92147 105689: contig of 13543 bp in length  
\* 105690 105789: gap of 100 bp  
\* 105790 117433: contig of 11644 bp in length  
\* 117434 117533: gap of 100 bp  
\* 117534 130091: contig of 12558 bp in length  
\* 130092 130191: gap of 100 bp  
\* 130192 149235: contig of 19044 bp in length.  
\* 149236 149235: Location/Qualifiers  
1. 149235  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-146F22"  
/clone\_id="RP11-146F22"  
1. 2046  
/note="assembly\_fragment"  
2147..2522  
/note="assembly\_fragment"  
clone end: T7  
vector\_side: right  
2623..4106  
/note="assembly\_fragment"  
4207..6021  
/note="assembly\_fragment"  
6122..8689  
/note="assembly\_fragment"  
8790..11605  
/note="assembly\_fragment"  
11706..15022  
/note="assembly\_fragment"  
15123..18721  
/note="assembly\_fragment"  
18822..23727  
/note="assembly\_fragment"  
23828..29512  
/note="assembly\_fragment"  
29613..34992  
/note="assembly\_fragment"  
35093..38620  
/note="assembly\_fragment"  
vector\_side: right  
38721..43918

misc\_feature

```

misc_feature /note="assembly_fragment"
44019..49427
/note="assembly_fragment"
misc_feature 49528..55671
/note="assembly_fragment"
misc_feature 55772..61066
/note="assembly_fragment"
misc_feature 61167..67660
/note="assembly_fragment"
misc_feature 67761..73637
/note="assembly_fragment"
misc_feature 73738..81886
/note="assembly_fragment"
misc_feature 81987..92046
/note="assembly_fragment"
misc_feature 92147..105689
/note="assembly_fragment"
misc_feature 105790..117433
/note="assembly_fragment"
misc_feature 117534..130091

```

Query Match 2.8%; Score 71; DB 2; Length 149235;  
 Best Local Similarity 100.0%; Pred. No. 12e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGACCTGATGATCCACCTTGGCTCCCAAGTCTGGATTACAG 2109  
 DB 49134 TCGAAGCTCTGACCTGATGATCCACCTTGGCTCCCAAGTCTGGATTACAG 49193

QY 2110 GTGTGAGCCAC 2120  
 DB 49194 GTGTGAGCCAC 49204

RESULT 41  
 HS1177E19/c 151188 bp DNA linear PRI 23-NOV-1999  
 LOCUS Human DNA sequence from clone 1177E19 on chromosome 1p36.12-36.31.  
 DEFINITION Contains the 3' part of the DNA-binding Zinc finger protein RIZ  
 gene, ESTE, an STS, GSSs and a CpG island, complete sequence.

ACCESSION AL031277.1 GI:4375907  
 VERSION HTG; CpG island; RIZ; zinc finger.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 151188)  
 AUTHORS Bagguley, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-FEB-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humqueyes@sanger.ac.uk; Clone  
 requests: clonerequests@sanger.ac.uk

## COMMENT

On Mar 7, 1999 this sequence version replaced gi:1153927.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 only a small overlap as described above.  
 This sequence is the entire insert of clone 1177E19. The true right  
 end of clone 726F20 (AL031273) is at 37052 in this sequence. This  
 sequence has been finished according to sequence map criteria as  
 follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome 1, constructed by the Sanger Centre Chromosome 1  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr1>  
 1177E19 is from the library RCTIS constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further  
 details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.  
 Location/Qualifiers  
 1..151188  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="p36.12-36.31"  
 /clone="RP5-117E19"  
 /clone\_1fb="RP1-5"  
 1..316  
 /note="LIPAI3 repeat: matches 5826..6156 of consensus"

```

repeat_region 316..387
/note="12 copies 6 mer gaaga 75% conserved"
repeat_region 317..390
/note="37 copies 2 mer ag 76% conserved"
repeat_region 400..450
/note="MIR repeat: matches 204..256 of consensus"
repeat_region 1157..1537
/note="LIR16A repeat: matches 25..440 of consensus"
repeat_region 2143..2937
/note="12 repeat: matches 1024..2712 of consensus"
repeat_region 2946..3419
/note="MTRIC repeat: matches 6..466 of consensus"
repeat_region 4592..4898
/note="Alusq repeat: matches 1..307 of consensus"
repeat_region 4917..5014
/note="LIMB8 repeat: matches 6070..6171 of consensus"
repeat_region 5910..6107
/note="Aluy repeat: matches 105..302 of consensus"
misc_feature complement(6172..6532)
/note="match: GSS A0053455"
repeat_region 7283..7573
/note="Alub repeat: matches 1..294 of consensus"
repeat_region 7641..7931
/note="Alusq repeat: matches 3..294 of consensus"
repeat_region 8682..8878
/note="MIR repeat: matches 11..247 of consensus"
repeat_region 9026..9042
/note="HAL1 repeat: matches 240..255 of consensus"
repeat_region 9043..9342
/note="Alusq repeat: matches 1..299 of consensus"
repeat_region 9343..9505
/note="HAL1 repeat: matches 255..431 of consensus"
repeat_region 9562..10370
/note="HAL1 repeat: matches 517..1188 of consensus"
repeat_region 10418..10515
/note="MIR2 repeat: matches 2..99 of consensus"
repeat_region 10638..10822
/note="HAL1 repeat: matches 1506..1707 of consensus"
repeat_region 10923..11142
/note="MIR repeat: matches 13..245 of consensus"
misc_feature 10952..11352
/note="match: GSS A000754"
repeat_region 11227..11466
/note="MTR1 repeat: matches 1..245 of consensus"
repeat_region 11493..13283
/note="LIPAI4 repeat: matches 4106..5885 of consensus"
repeat_region 13284..13628
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 13629..15238
/note="THEIC-internal repeat: matches 3..1580 of consensus"
repeat_region 15241..15600
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 15601..15865
/note="LIPAI4 repeat: matches 5881..6149 of consensus"
repeat_region 15922..16083
/note="LIMC/D repeat: matches 5311..5477 of consensus"
repeat_region 16120..16369
/note="LIMC repeat: matches 1756..1774 of consensus"
repeat_region 16370..16885
/note="MTR1D repeat: matches 1..505 of consensus"

```

repeat_region	16886..17391	/note="LM6C repeat: matches 1713..2197 of consensus"
repeat_region	17407..17588	/note="LM7A repeat: matches 2433..2617 of consensus"
repeat_region	17593..17920	/note="LM4 repeat: matches 17593..17920 of consensus"
repeat_region	17921..18278	/note="THE1C repeat: matches 46..371 of consensus"
repeat_region	18280..18589	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	18590..19232	/note="THE1B-INTERNAL repeat: matches 1269..1579 of consensus"
repeat_region	19059..19232	/note="MER6 repeat: matches -64..838 of consensus"
repeat_region	19233..20511	/note="THE1B-INTERNAL repeat: matches 5..1284 of consensus"
repeat_region	20516..20871	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	20872..20926	/note="MST1 repeat: matches 1..55 of consensus"
repeat_region	20927..22462	/note="MST1-internal repeat: matches 1..1580 of consensus"
repeat_region	22463..22839	/note="THE1C repeat: matches 1..371 of consensus"
repeat_region	22840..24495	/note="LM4 repeat: matches 2620..4287 of consensus"
repeat_region	24496..24790	/note="AL89 repeat: matches 1..292 of consensus"
repeat_region	24791..25222	/note="LM4 repeat: matches 4287..4716 of consensus"
repeat_region	25246..25474	/note="LM1A12 repeat: matches -1415..-1183 of consensus"
repeat_region	25497..25963	/note="LM22 repeat: matches 1..505 of consensus"
repeat_region	25974..26086	/note="LIP repeat: matches 1703..1815 of consensus"
repeat_region	26087..27013	/note="LM1A10 repeat: matches 2626..3615 of consensus"
repeat_region	27014..27082	/note="THE1B repeat: matches 1..67 of consensus"
repeat_region	27122..27413	/note="THE1C repeat: matches 80..371 of consensus"
repeat_region	27436..28294	/note="LM1A10 repeat: matches 3644..5531 of consensus"
repeat_region	28317..28946	/note="5 copies 6 mer ttatata 100% conserved"
repeat_region	29348..29767	/note="LM1A4 repeat: matches 5727..6146 of consensus"
repeat_region	30002..30195	/note="MIR repeat: matches 58..261 of consensus"
repeat_region	30196..30329	/note="MIR repeat: matches 58..261 of consensus"
repeat_region	30330..30374	/note="LM1A10 repeat: matches 5514..6158 of consensus"
repeat_region	30976..31286	/note="MIR repeat: matches 1..311 of consensus"
repeat_region	31324..31626	/note="MIR repeat: matches 1..308 of consensus"
repeat_region	31743..32896	/note="LM1A10 repeat: matches 4734..5889 of consensus"
repeat_region	32897..33313	/note="MIR repeat: matches 1..303 of consensus"
repeat_region	33314..33513	/note="MIR repeat: matches 1..303 of consensus"
repeat_region	33525..34025	/note="match: GSS AQ132604"
repeat_region	33828..33989	/note="MIR repeat: matches 241..409 of consensus"
repeat_region	34165..34479	/note="MIR repeat: matches 8..308 of consensus"
repeat_region	35301..35771	/note="MIR repeat: matches 1..504 of consensus"
repeat_region	36083..36193	/note="MIR repeat: matches 1..504 of consensus"

	repeat_region	/note="AlusX repeat: matches 1. .308 of consensus" 37664. .37951
	repeat_region	/note="MIR repeat: matches 84. .173 of consensus" 38243. .38397
	repeat_region	/note="MIR repeat: matches 61. .227 of consensus" 38453. .38752
	repeat_region	/note="AlusG repeat: matches 1. .301 of consensus" 39596. .39774
	repeat_region	/note="MERF3 repeat: matches 1. .182 of consensus" 40042. .40500
	repeat_region	/note="L2 repeat: matches 2150. .2711 of consensus" 40546. .40636
	repeat_region	/note="MER96 repeat: matches 85. .175 of consensus" 41856. .41997
	repeat_region	/note="MER54 repeat: matches 40. .186 of consensus" 42015. .42168
	repeat_region	/note="MIR repeat: matches 34. .199 of consensus"
Query Match		2.8%; Score 71; DB 9; Length 151189;
Best Local Similarity		100.0%; Pred. No. 1.2e-28;
Matches	71; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	2050 TCGAAGCTGTACCTCAGGTGATTCACCCACCCTTGGCTCCCAAGAAGTGCTGGATTACAG	2109
Db	38534 TCGAACCTCGAACCTCAGGTGATTCACCCACCCTTGGCTCCCAAGAAGTGCTGGATTACAG	38475
Oy	2110 GTGTGAGCCAC	2120
Db	38474 GTGTGAGCCAC	38464

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
42	AC026110/c	Homo sapiens 12 BAC RPL1-43433 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	AC026110	AC026110	HTG.	human.	Homo sapiens
		153190 bp DNA linear PRI 01-JUN-2002					
REFERENCE							Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS							1 (bases 1 to 153190)
							Mazny,D.M., Adams,C., Adio-Ogunola,B., Ali-osman,F.R., Allen,C., Alebrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Birnagge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chin,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Cuyler,M.D., Dathorne,S.R., David,R., Davila,M., Davis,C., Day-Carroll,L., Dedereich,D.A., Delaney,K.R., Delgado,O., Demm,A.L., Ding,Y., Dinh,H.H., Douchevaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Fails,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R., Garrett,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honski,F., Howard,S., Huber,J., Hulik,S., Hume,J., Ioshkhesi,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Jordan,S., Kalsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureish,A., Landty,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lottesgud,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondei,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawlinley,E., Mcloed,M.P., Meador,M., Mel,G., Mescher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,N., Morris,S., Moser,W., Neal,D., Nelson,D., Newton,J., Newton,M., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemko,S.,

	Gyuh,M., Okunubu,G., Osegunye,N., Owiedo,R., Pace,A., Paul,B., Peery,J., Peters,L., Pickens,R., Primus,E., Pullen,T., Qureshi,S., Ren,Y., Rivers,M., Rojas,A., Rojchokan,I., Rolle,M., Ruiz,C., Savory,G., Scherer,S., Scott,G., Shen,H., Shum,C., Shoostarian,I., Stason,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tameitsa,A., Tameitsa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telifod,B., Thomas,N., Thomas,S., Uemami,K., Vaequez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,X.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstein,G. and Gibbs.R.
JOURNAL TITLE	Direct Submission Unpublished
AUTHORS	2 (bases 1 to 153190)
TITLE	Worley,K.C. Direct Submission
JOURNAL	Submitted (19-Mar-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 153190) Worley,K.C. Direct Submision
AUTHORS	Submitted (03-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
TITLE	4 (bases 1 to 153190) Worley,K.C. Direct Submission
REFERENCE	Submitted (01-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	On Jul 3, 2001 this sequence version replaced gi:1454364. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-hejp@bcm.tmc.edu
COMMENT	CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.
	ANNOTATION OF FEATURES:  STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.  Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similariaty (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintaining sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.
	SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.
	QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <a href="http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html">http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html</a> .
FEATURES SOURCE	. QUALSTAT-REPORT Location/Qualifiers . . 1..153190

repeat_region	/organism="Homo sapiens"
STS	/db_xref="taxon:9606"
STS	/chromosome="12"
STS	/clone="RP11-434E3"
repeat_region	28. .71
repeat_region	/rpt_family="A-rich"
repeat_region	113. .317
repeat_region	/standard_name="164818"
repeat_region	240. .319
repeat_region	/standard_name="164799"
repeat_region	complement(432. .746)
repeat_region	/rpt_family="AluSk"
repeat_region	1106. .1205
repeat_region	/rpt_family="HERV16"
repeat_region	1207. .1655
repeat_region	/rpt_family="MEL1C"
repeat_region	1885. .2168
repeat_region	/rpt_family="AluJo"
repeat_region	2165. .2199
repeat_region	/rpt_family="7AAA)n"
repeat_region	2897. .3666
repeat_region	/rpt_family="HERV16"
repeat_region	3667. .3689
repeat_region	/rpt_family="(A)n"
repeat_region	3851. .4586
repeat_region	/rpt_family="HERVL"
repeat_region	4709. .5667
repeat_region	/rpt_family="MERVL"
repeat_region	complement(5666. .5978)
repeat_region	/rpt_family="AluSk"
repeat_region	5979. .6090
repeat_region	/rpt_family="MERVL"
repeat_region	6091. .6368
repeat_region	/rpt_family="AluSk"
repeat_region	6369. .6890
repeat_region	/rpt_family="MERVL"
repeat_region	7208. .7360
repeat_region	/rpt_family="FLAM_C"
repeat_region	7583. .7746
repeat_region	/rpt_family="LTR16A1"
repeat_region	complement(7747. .8057)
repeat_region	/rpt_family="AluSk"
repeat_region	8058. .8179
repeat_region	/rpt_family="LTR16A1"
repeat_region	complement(8184. .8659)
repeat_region	/rpt_family="L2"
repeat_region	9007. .9129
repeat_region	/rpt_family="MIR"
repeat_region	9130. .9297
repeat_region	/rpt_family="FRAM"
repeat_region	complement(10580. .10877)
repeat_region	/rpt_family="AluJo"
repeat_region	11607. .11636
repeat_region	/rpt_family="AT-rich"
repeat_region	13245. .13288
repeat_region	/rpt_family="L2"
repeat_region	complement(13336. .13567)
repeat_region	/rpt_family="L1PA8"
repeat_region	13568. .13614
repeat_region	/rpt_family="L1PA"
repeat_region	13661. .13894
repeat_region	/rpt_family="L2"
repeat_region	13884. .13941
repeat_region	/rpt_family="L2"
repeat_region	complement(14227. .14442)
repeat_region	/rpt_family="MIR"
repeat_region	14449. .14556
repeat_region	/rpt_family="(GC)n"
repeat_region	complement(14537. .14567)
repeat_region	/rpt_family="L1M8"
repeat_region	complement(15591. .15636)
repeat_region	/rpt_family="MIR"
repeat_region	complement(15713. .16028)



```

repeat_region      /rpt_family="12"
16163 .16588
/rpt_family="11P13"
repeat_region      /rpt_family="11P13"
17097 .17130
/rpt_family="11P13"
repeat_region      /rpt_family="11P13"
17131 .17423
/rpt_family="11P13"
repeat_region      /rpt_family="11P13"
17915 .17946
/rpt_family="11P13"
repeat_region      /rpt_family="11P13"
18738 .18865
/rpt_family="11P13"
repeat_region      /rpt_family="11P13"
19181 .19273
/rpt_family="11P13"
repeat_region      /rpt_family="11P13"
20066 .20328
/rpt_family="11P13"

```

```

Query Match      2.0% Score 71 DB 9 Length 153190;
Best Local Similarity 100.0%; Pred. No. 1.2e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2050 TCGAAGCTGACCTGAGTATCCACCCACCTTGCTCCCAAGTCTGGAGTTACAG 2109
Db 132311 TCGAAGCTGACCTGAGTATCCACCCACCTTGCTCCCAAGTCTGGAGTTACAG 132252

```

```

QY 2110 GTGTGAGCCAC 2120
Db 132251 GTGTGAGCCAC 132241

```

```

RESULT 43
AL513012/c 15353 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-26602, *** SEQUENCING IN
DEFINITION PROGRESS ***
ACCESSION AL513012.4 GI:12733808
VERSION HTG; HTGS_PHASE1; HTGS_CANCELLED.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 15353)
McLay, K.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requesters: clonerequest@sanger.ac.uk
On Feb 9, 2001 this sequence version replaced gi:12584494.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba26802
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator ABI; 1% of reads
Chemistry: Dye-terminator ET-amersham; 6% of reads
Dye-terminator Big Dye; 92% of reads
Consensus quality: 140250 bases at least Q40
Consensus quality: 145215 bases at least Q30
Consensus quality: 148460 bases at least Q20
Insert size: 151053; sum-of-coverage
Insert size: 200511; 5.3% error; agarose-fp
Quality coverage: 5.07x in Q20 bases; sum-of-coverage
Quality coverage: 4.44x in Q20 bases; agarose-fp

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2078: contig of 2078 bp in length
2079 2178: gap of 100 bp
2179 10085: contig of 7907 bp in length
10086 10185: gap of 100 bp
10186 12349: contig of 2164 bp in length
12350 12449: gap of 100 bp
12450 125415: contig of 12966 bp in length
125416 25519: gap of 104 bp
25520 26046: contig of 2527 bp in length
26047 28146: gap of 100 bp
28147 30184: contig of 2038 bp in length
30185 30284: gap of 100 bp
30285 62042: contig of 31758 bp in length
62043 62142: gap of 100 bp
62143 69938: contig of 7696 bp in length
69939 69938: gap of 100 bp
69939 72212: contig of 2274 bp in length
72213 72212: gap of 100 bp
72213 74404: contig of 2092 bp in length
74405 74504: gap of 100 bp
74505 78593: contig of 4089 bp in length
78594 78693: gap of 100 bp
78694 81831: contig of 3138 bp in length
81832 81931: gap of 100 bp
81932 84050: contig of 2119 bp in length
84051 84150: gap of 100 bp
84151 88043: contig of 3893 bp in length
88044 88143: gap of 100 bp
88144 93493: contig of 5350 bp in length
93494 93593: gap of 100 bp
93594 101833: contig of 8240 bp in length
101834 101933: gap of 100 bp
101934 104335: contig of 2402 bp in length
104336 104435: gap of 100 bp
104436 107388: contig of 2953 bp in length
107389 107488: gap of 100 bp
107489 132346: contig of 24858 bp in length
132347 132446: gap of 100 bp
132447 137068: contig of 4632 bp in length
137069 137168: gap of 100 bp
137169 139335: contig of 2167 bp in length
139336 139435: gap of 100 bp
139436 142713: contig of 3278 bp in length
142714 142813: gap of 100 bp
142814 145741: contig of 2928 bp in length
145742 145841: gap of 100 bp
145842 148055: contig of 2214 bp in length
148056 148155: gap of 100 bp
148156 150735: contig of 2580 bp in length
150736 150835: gap of 100 bp
150836 153533: contig of 2718 bp in length.
Location/Qualifiers
1. 153533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-26602"
/clone_id="RP11-1.1"
1. 2078
/contig="assembly_fragment:00410"
fragment_chain:1
2179. 10085
/contig="assembly_fragment:00093"
fragment_chain:1
10186. 12349
/contig="assembly_fragment:01777"
fragment_chain:1
12450. 25415
/contig="assembly_fragment:00225"
fragment_chain:1

```

```

misc_feature
10186. 12349
misc_feature
12450. 25415
misc_feature
150836. 153533

```



KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

HTG.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 155923)  
Birren, B., Nusbaum, C., and Lander, E.  
Homo sapiens chromosome 17, clone RP11-426D19  
unpublished  
2 (bases 1 to 155923)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,  
Camata, J., Campolano, A., Chang, J., Choepel, Y., Colangelo, M.,  
Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K.,  
Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,  
Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Margus, N., Mathews, C., McCarthy, M.,  
McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Menus, L.,  
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Ribbeck, M., Riley, R., Rise, C., Rogov, P., Roman, S.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,  
Severy, P., Sougnier, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S.,  
Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 155923)  
Birren, B., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A.,  
Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C.,  
McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 155923)  
Birren, B., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A.,  
Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C.,  
McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (18-JUL-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 18, 2002 this sequence version replaced gi:21699272.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L13059  
Center clone name: 426\_D\_19  
-----  
SFE end overlaps AC015883 [WICGR L725], 17 end overlaps AC003950  
[WICGR project L261]. We are submitting the first 155,9 kilobases  
of this project.  
----- Location/Qualifiers  
1.155923  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-426D19"  
/clone\_lib="RP11-426D19 Human Male BAC"  
727.731  
/note="<30 qual SNGL region"  
repeat\_region  
991.1039  
/rpt\_family="(GAAA)n"  
1463.1497  
/rpt\_family="(A)n"  
1605.1609  
/note="<30 qual SNGL region"  
1662.1666  
/note="<30 qual SNGL region"  
1957.1961  
/note="<30 qual SNGL region"  
1979.1983  
/note="<30 qual SNGL region"  
2263.2397  
/rpt\_family="AluSx"  
2398.2706  
/rpt\_family="AluSg"  
2707.2842  
/rpt\_family="AluSx"  
complement(2861..3165)  
/rpt\_family="AluY"  
3740.4036  
/rpt\_family="AluY"  
4558.4666  
/rpt\_family="MIR3"  
complement(6158..6455)  
/rpt\_family="AluSx"  
7397.7503  
/rpt\_family="LTR67"  
7553.7723  
/rpt\_family="MIR3"  
complement(7734..7992)  
/rpt\_family="AluSx"  
8405.8511  
/rpt\_family="L1ME4A"  
8973.9024  
/rpt\_family="AT-rich"  
9085.9289  
/rpt\_family="L1ME4A"  
9388.9408  
/rpt\_family="(TG)n"  
9456.9581  
repeat\_region

Query Match	Best Local Similarity	Conservative	Mismatches	Indels	Gaps
2050	TCGAACCTCTGACCTCAGGTATCCACCACTTGCGCTCCCAAGTCTGGGATTACG	2109			
58944	TCGAACCTCTGACCTCAGGTATCCACCACTTGCGCTCCCAAGTCTGGGATTACG	58885			
2110	GTGTGAGCCAC	2120			
58884	GTGTGAGCCAC	58874			
RESULT 46	AC073839/c	15824 bp	DNA	linear	HTG 15-JUL-2000
LOCUS	AC073839/c	15824 bp	DNA	linear	HTG 15-JUL-2000
DEFINITION	Homo sapiens chromosome 17 clone RP11-426D19, WORKING DRAFT				
SEQUENCE	6 unordered pieces.				
ACCESSION	AC073839				
VERSION	AC073839.1				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens.				

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 158524)
JOURNAL	Waterston,R.H.
REFERENCE	2 (bases 1 to 158524)
AUTHORS	Waterston,R.H.
TITLE	Submitted (29-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_MH0426D19
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.9901319
Consensus quality: 155368 bases at least Q40
Consensus quality: 156306 bases at least Q30
Consensus quality: 156858 bases at least Q20
Insert size: 197000; agarose-1p
Insert size: 158024; sum-of-contigs
Quality coverage: 4.95 in Q20 bases; agarose-1p
Quality coverage: 6.21 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3334: contig of 3334 bp in length
* 3335 3344: gap of unknown length
* 3435 23416: contig of 19982 bp in length
* 23417 23516: gap of unknown length
* 23517 50419: contig of 26903 bp in length
* 50420 50519: gap of unknown length
* 50520 79336: contig of 28807 bp in length
* 79337 79436: gap of unknown length
* 79437 113976: contig of 34550 bp in length
* 113977 114077: gap of unknown length
* 114077 158524: contig of 44448 bp in length.
*
* Location/Qualifiers
*
* 1. 158524
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="17"
* /clone="RP11-426D19"
*
* 1. 3334
* /note="assembly_name:Contig4"
* 3435 23416
* /note="assembly_name:Contig5"
* 23517..50419
* /note="assembly_name:Contig6"
* 50520..79336
* /note="assembly_name:Contig7"
* 79427..113976
* /note="assembly_name:Contig8
* clone_end:77
* vector_side:left"
* 114077..158524
* /note="assembly_name:Contig9

```

clone\_end:SP6  
vector\_side:right"  
BASE COUNT 43403 a 35269 c 35618 g 43713 t 521 others  
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 158524;  
Best Local Similarity 100.0%; Pred. No. 1,1e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAATCTTGACCTGAGGTGATCCACCCCTTGCCCTCCCAAGTCTGGGATTACAG 2109  
Db 61678 TCGAATCTTGACCTGAGGTGATCCACCCCTTGCCCTCCCAAGTCTGGGATTACAG 61619

Qy 2110 GTGTGAGCCAC 2120  
Db 61618 GTGTGAGCCAC 61608

RESULT 47  
AC025145 161059 bp DNA linear HTG 21-MAR-2000  
LOCUS Homo sapiens chromosome 1 clone RP11-26P17 map 1, WORKING DRAFT  
DEFINITION  
SEQUENCE 25 unordered pieces.  
AC025145  
AC025145.2 GI:7272268  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
AUTHORS Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Tittle JOURNAL (bases 1 to 161059)  
REFERENCE  
AUTHORS Biren, B., Linton, L., Nusbaum, C., and Lander, E.  
JOURNAL (bases 1 to 161059)  
REFERENCE  
AUTHORS Homo sapiens chromosome 1, clone RP11-26P17  
unpublished

2 (bases 1 to 161059)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barina, N., Bastien, V., Bede, F.,  
Boguslavsky, L., Bouckgalter, B., Brown, A., Buckett, G.,  
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hages, B., Heald, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehotzky, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Mejdrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisanti, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,  
Teifaye, S., Theodore, J., Tittel, A., Travers, M., Triggillo, J.,  
Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA.  
On Mar 21, 2000 this sequence version replaced gi:715896.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W1BR  
Web site: http://www-seg.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L7960  
Center clone name: 26\_P\_17  
Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads

## FEATURES

## source

Chemistry: Dye-terminator Big Dye, 100% of reads  
Assembly program: Phrap; version 0.960721  
Consensus quality: 146317 bases at least Q40  
Consensus quality: 154085 bases at least Q30  
Consensus quality: 156792 bases at least Q20  
Insert size: 16000; agarose-fp  
Insert size: 15859; sum-of-contigs  
Quality coverage: 4.4 in Q20 bases; agarose-fp  
Quality coverage: 4.6 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1146: contig of 1146 bp in length  
1147 1246: gap of 100 bp  
1247 2748: contig of 1502 bp in length  
2749 2848: gap of 100 bp  
2849 4365: contig of 1517 bp in length  
4366 4465: gap of 100 bp  
4466 5524: contig of 1059 bp in length  
5525 5624: gap of 100 bp  
5625 7199: contig of 1575 bp in length  
7200 7299: gap of 100 bp  
7300 9128: contig of 1829 bp in length  
9129 9228: gap of 100 bp  
9229 10151: contig of 923 bp in length  
10152 10251: gap of 100 bp  
10252 13233: contig of 2982 bp in length  
13234 13333: gap of 100 bp  
13334 17255: contig of 3922 bp in length  
17256 17355: gap of 100 bp  
17356 19626: contig of 2271 bp in length  
19627 19726: gap of 100 bp  
19727 21287: contig of 1561 bp in length  
21288 21387: gap of 100 bp  
21388 25830: contig of 4443 bp in length  
25831 25930: gap of 100 bp  
25931 30523: contig of 4593 bp in length  
30524 30623: gap of 100 bp  
30624 40584: contig of 9961 bp in length  
40585 40684: gap of 100 bp  
40685 48850: contig of 8166 bp in length  
48851 48950: gap of 100 bp  
48951 54874: contig of 5924 bp in length  
54875 54974: gap of 100 bp  
54975 66089: contig of 11115 bp in length  
66090 66189: gap of 100 bp  
66190 73670: contig of 7481 bp in length  
73671 73770: gap of 100 bp  
73771 83280: contig of 9510 bp in length  
83281 83380: gap of 100 bp  
83381 92734: contig of 9354 bp in length  
92735 92834: gap of 100 bp  
92835 102275: contig of 9441 bp in length  
102276 102375: gap of 100 bp  
102376 114577: contig of 12202 bp in length  
114578 114677: gap of 100 bp  
114678 129572: contig of 14895 bp in length  
129573 129672: gap of 100 bp  
129673 143684: contig of 14012 bp in length  
143685 143784: gap of 100 bp  
143785 161059: contig of 17275 bp in length.  
Location/Qualifiers  
1. 161059  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"

misc\_feature /clone="RP11-26P17"  
 misc\_feature /clone.lib="RP11-11 Human Male BAC"  
 misc\_feature 1..1146  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature 1247..2748  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature 2849..4365  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature 4466..5524  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature 5625..7199  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature 7300..9128  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature 9229..10151  
 misc\_feature /note="assembly\_fragment"

misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"

misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"

misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"

misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"

misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"

misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"  
 misc\_feature /note="assembly\_fragment"

BASE COUNT 42532 a 37330 c 38375 g 40411 t 2407 others  
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 161059;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCGAGCTGAGTGCACCACTTGGCTCCCAAGTGTGGAATTACAG 2109  
 Db 79231 TCGAAGCTCGAGCTGAGTGCACCACTTGGCTCCCAAGTGTGGAATTACAG 79290  
 QY 2110 GTGTGAGCCAC 2120  
 Db 79291 GTGTGAGCCAC 79301

RESULT 48

AC025145/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 161059)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 1, clone RP11-26P17  
 Unpublished  
 2 (bases 1 to 161059)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,  
 Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karavas,A.,  
 Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,  
 Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Homann,N., Stojanovic,N., Sudraman,A., Taimas,J.,  
 Testaye,S., Theodore,J., Tittell,A., Travers,M., Triggillo,U.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission

TITLE  
 JOURNAL  
 COMMENT  
 Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research 370 Charles Street, Cambridge, MA 02141, USA  
 On Mar 21, 2000 this sequence version replaced gi:1158956.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center Project name: U7960  
 Center Clone name: 26\_P\_17

Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 146317 bases at least Q40  
 Consensus quality: 154095 bases at least Q30  
 Consensus quality: 156792 bases at least Q20  
 Insert size: 166000; agarose-fp  
 Insert size: 158659; sum-of-coverage  
 Quality coverage: 4.4 in Q20 bases; sum-of-coverage  
 Quality coverage: 4.6 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently  
 consists of 25 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1 1146: contig of 1146 bp in length



TITLE  
JOURNAL

## COMMENT

Coke, P., Deatellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,  
Galligan, J., Gardyna, S., Grant, G., Hages, B., Heaford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karcas, A., Klein, J.,  
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
McEwen, P., McGurk, A., McKernan, K., McDonald, J., Meidrim, J.,  
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testaye, S., Turrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (02-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 4, 2000 this sequence version replaced gi:1107860.

All repeats were identified using RepeatMasker:

http://ftp.genome.washington.edu/RM/RepeatMasker.html

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W18

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

## Project Information

Center project name: L2709

Center clone name: 3\_C\_18

## Sequencing Vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158604 bases at least Q40

Consensus quality: 161244 bases at least Q30

Consensus quality: 162123 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 162621; sum-of-contigs

Quality coverage: 6.1 in Q20 bases; agarose-fp

Quality coverage: 6.2 in Q20 bases; sum-of-contigs

\*\*\*\*\*

NOTE: This is a 'working draft' sequence. It currently  
consists of 9 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 3528: contig of 3528 bp in length  
3529 3628: gap of 100 bp  
3629 6379: contig of 2751 bp in length  
6380 6479: gap of 100 bp  
6480 11952: contig of 5473 bp in length  
11953 12052: gap of 100 bp  
12053 87340: contig of 75288 bp in length  
87341 87440: gap of 100 bp  
87441 93389: contig of 5949 bp in length  
93390 93489: gap of 100 bp  
93490 107761: contig of 14272 bp in length  
107762 107861: gap of 100 bp  
107862 125525: contig of 17664 bp in length  
125526 125625: gap of 100 bp  
125626 160209: contig of 34584 bp in length  
160210 160309: gap of 100 bp  
160310 163421: contig of 3112 bp in length.

## FEATURES

## source

1. 163421

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="RP11-3C18"

/clone\_id="RP11 Human Male BAC"

## misc\_feature

1. 3528

/note="assembly\_fragment"

clone end:SP6

vector\_side:left"

misc\_feature 3629..6379 /note="assembly\_fragment"  
misc\_feature 6480..11952 /note="assembly\_fragment"  
misc\_feature 12053..87340 /note="assembly\_fragment"  
misc\_feature 87441..93389 /note="assembly\_fragment"  
misc\_feature 93490..107761 /note="assembly\_fragment"  
misc\_feature 107862..125525 /note="assembly\_fragment"  
misc\_feature 125626..160209 /note="assembly\_fragment"  
misc\_feature 160310..163421 /note="assembly\_fragment"  
misc\_feature clone\_end:T7 vector\_side:right"  
BASE COUNT 47341 a 35177 c 33775 g 46328 t 800 others  
ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 163421;  
Best Local Similarity 100.0%; Pred. No. 1.1e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2050 TCGAAGCTCTGACTGATGATTCACCCACCTTGCTCCCAAGTCTGGATTACG 2109  
DB 35147 TCGAAGCTCTGACTGATGATTCACCCACCTTGCTCCCAAGTCTGGATTACG 35206  
OY 2110 GTGTGAGCCAC 2120  
DB 35207 GTGTGAGCCAC 35217

RESULT 50  
AL158137  
LOCUS 163816 bp DNA linear HTG 10-JUL-2001  
DEFINITION Homo sapiens chromosome 6 clone RP1-506112, \*\*\* SEQUENCING IN  
PROCESSES \*\*\* 4 unordered pieces.  
ACCESSION AL158137  
VERSION AL158137.4 GI:11875671  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Carniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 163816)  
REFERENCE  
AUTHORS Almeida, J.  
TITLE Direct Submission  
JOURNAL CBIO 15A, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Dec 17, 2000 this sequence version replaced gi:9931671.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: d1506112

----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 162746 bases at least Q40  
Consensus quality: 162931 bases at least Q30  
Consensus quality: 163148 bases at least Q20  
Insert size: 163516; sum-of-contigs  
Insert size: 144035; 25.1% error; agarose-fp  
Quality coverage: 7.99% in Q20 bases; sum-of-contigs Quality  
coverage: 9.07% in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently



\* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 25098: contig of 25098 bp in length  
 \* 25099 25198: gap of 100 bp  
 \* 25199 100746: contig of 75548 bp in length  
 \* 100747 100846: gap of 100 bp  
 \* 100847 146855: contig of 46009 bp in length  
 \* 146856 146955: gap of 100 bp  
 \* 146956 163816: contig of 16861 bp in length.

## FEATURES

source  
 1.163816  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP3-506112"  
 /clone\_lib="RPC1-3"  
 1.25098  
 /note="assembly fragment:01964  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left"  
 25199.100746  
 /note="assembly fragment:00823  
 fragment\_chain:1"  
 100847.146855  
 /note="assembly fragment:01285  
 fragment\_chain:1"  
 146956.163816  
 /note="assembly fragment:00518  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 50468 a 34486 c 33411 g 45142 t 309 others  
 ORIGIN

Query Match 2.8%; Score 71; DB 2; Length 163816;  
 Best Local Similarity 100.0%; Pred. NO. 1.1e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGAGCTCAGGTGATCCACCCAGCTTGGCTCCCAAGTGTGGATTACAG 2109  
 Db 24897 TCGAAGCTCTGAGCTCAGGTGATCCACCCAGCTTGGCTCCCAAGTGTGGATTACAG 24956  
 QY 2110 GTGTGAGCCAC 2120  
 Db 24957 GTGTGAGCCAC 24967

RESULT 51  
 HS120G22 166518 bp DNA linear PRI 21-JUL-2000  
 LOCUS Human DNA sequence from clone Rpl-120G22 on chromosome  
 DEFINITION 1p36.21-36.33, complete sequence.  
 ACCESSION AL031847  
 VERSION AL031847.17 GI:9369286  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 166518)  
 REFERENCE Wray, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Jul 22, 2000 this sequence version replaced gi:9368784.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Info, information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Ch1  
 Rpl-120G22 is from the library RPC1-1 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see http://bacpac.med.buffalo.edu/  
 VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone  
 Rpl-120G22. It may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true right end of clone Rpl-120G22 is at 166518 in this  
 sequence. The true left end of clone Rps-889923 is at 90508 in this  
 sequence. The true right end of clone Rpl-20208 is at 100 in this  
 sequence.

## FEATURES

source  
 1.166518  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="p36.21-36.33"  
 /clone="Rpl-120G22"  
 /clone\_lib="RPC1-1"  
 8814.9131  
 /note="match: GSS: Em:AQ716455"  
 8816.9229  
 /note="match: GSS: Em:AQ239786"  
 10012.10541  
 /note="match: GSS: Em:AQ678815"  
 18389.18823  
 /note="match: GSS: Em:AQ270793"  
 25588.26045  
 /note="match: GSS: Em:AQ214655"  
 25694.25967  
 /note="match: GSS: Em:AQ097494"  
 complement(41087.41505)  
 /note="match: GSS: Em:AQ213363"  
 complement(41123.41505)  
 /note="match: GSS: Em:AQ133445"  
 45297.45591  
 /note="match: GSS: Em:AF046498"  
 49180.49534  
 /note="Single clone region. short insert library only"  
 complement(57981.58764)  
 /note="match: GSS: Em:AF158011"  
 complement(60582.61043)  
 /note="match: GSS: Em:AQ655221"  
 64797.64841  
 /note="Single clone region. short insert library only"  
 68250.68620  
 /note="match: STS: Em:G16281"  
 70460.70622  
 /note="match: STS: Em:G21080"  
 complement(74554.74895)  
 /note="match: GSS: Em:AQ097224"  
 complement(79470.79747)  
 /note="match: STS: Em:AF191963"  
 complement(79556.79740)

```

misc_feature /note="match: GSS: Em:AQ541496"
               complement(79935..80319)
               /note="match: STS: Em:HSAS2WG1"
               complement(85942..86372)
misc_feature /note="match: STS: Em:G22706"
               complement(86963..87322)
misc_feature /note="match: GSS: Em:AQ191756"
               complement(87339..87808)
misc_feature /note="match: GSS: Em:AQ698012"
               complement(87809..89285)
misc_feature /note="match: GSS: Em:AQ121434"
               complement(94030..94434)
misc_feature /note="match: GSS: Em:B89150"
               complement(94030..94405)
misc_feature /note="match: GSS: Em:AQ008338"
               complement(94054..94464)
misc_feature /note="match: GSS: Em:AQ010380"
               complement(112106..112211,116809..116933,123047..123132)
               /note="match: GSS: Em:AQ634365"
               complement(join(112121..112215,116809..116936))
               /note="match: GSS: Em:AQ202845"
               complement(112126..112224)
               /note="match: GSS: Em:AF046385"
               complement(join(112145..112211,116809..116936))
               /note="match: GSS: Em:AQ534370"
               complement(join(112150..112211,116809..116933,
               123047..123172))
               /note="match: GSS: Em:AQ489599"
               complement(113100..113712)
               /note="match: GSS: Em:AQ47048"
               complement(113175..113715)
               /note="match: GSS: Em:AQ76615"
               complement(113710..114090)
               /note="match: GSS: Em:AQ34236"
               complement(116809..116933,123047..123119)
               /note="match: GSS: Em:AQ32842"
               complement(116809..116941)
               /note="match: GSS: Em:AQ357462"
               complement(116809..116936)
               /note="match: GSS: Em:AQ047297 Em:AQ351276"
               complement(join(116840..116933,123047..123191))
               /note="match: STS: Em:G13224 Em:G13474"
               complement(121507..121762)
               /note="match: GSS: Em:AQ482227"
               complement(123035..123248)
               /note="match: GSS: Em:AQ779080"
               complement(123142..123248)
               /note="match: GSS: Em:AQ634365"
               complement(123192..123356)
               /note="match: STS: Em:G11177"
               complement(124453..124584)
               /note="match: GSS: Em:AQ002260"
               complement(124491..124832)
               /note="match: STS: Em:G06939"
               complement(129943..130018)
               /note="Single clone region"
               complement(14105..144499)
               /note="match: GSS: Em:AQ231959"
               complement(144105..144483)
               /note="match: GSS: Em:AQ231956"
               complement(148672..149104)
               /note="match: GSS: Em:AQ672446"
               complement(152031..152510)
               /note="match: GSS: Em:AQ233353"
               complement(152031..152510)
BASE COUNT 34311 a 45982 c 47124 g 39101 t
ORIGIN

```

```

Query Match      2.8% Score 71: DB 9: Length 166518;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY * 2050 TCGAAGCTGAGCTAGATGATCCACCCAGCTGCTCCCAAGTGTGGATTACAG 2109
|||||

```

```

Db 105225 TCGAAGCTGAGCTAGATGATCCACCCAGCTGCTCCCAAGTGTGGATTACAG 105284
QY 2110 GTGTGAGCCAC 2120
|||||
Db 105285 GTGTGAGCCAC 105295

```

```

RESULT 52
AL357078
LOCUS
DEFINITION
Homo sapiens chromosome 1 clone RP4-535B20, *** SEQUENCING IN
PROGRESS ***, 4 unordered pieces.
ACCESSION
AL357078
VERSION
AL357078.3 GI:22003096
KEYWORDS
HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 167055)
REFERENCE
Submitted (12-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, GB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone request: clonequest@sanger.ac.uk
On Jul 29, 2002 this sequence version replaced gi:21955468.
COMMENT
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: d0535820
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator; 14% of reads
Dye-terminator Big Dye; 76% of reads
Consensus quality: 15548 bases at least Q40
Consensus quality: 16106 bases at least Q30
Consensus quality: 16474 bases at least Q20
Insert size: 16675; sum-of-contigs
Insert size: 154002; 9.7% error; agarose-fp
Quality coverage: 8.38x in Q20 bases; sum-of-contigs Quality
coverage: 9.63x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 149929: contig of 149929 bp in length
* 149930 150029: gap of 100 bp
* 150030 161415: contig of 11386 bp in length
* 161416 161515: gap of 100 bp
* 161516 164717: contig of 3202 bp in length
* 164718 164817: gap of 100 bp
* 164818 167055: contig of 2238 bp in length.

```

## FEATURES

```

source
1..167055
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-535B20"
/clone_id="RPI-4"
1..149929
/note="assembly_fragment:03564
clone_end:17
vector_side:left"
misc_feature

```



[illegible]

	misc_feature	complement(45737..46222)	/note="match: GSS: Em:RQ39959"
	repeat_region	46482..46801	
	repeat_region	/note="ALUSK:repeat: matches 1..312 of consensus"	
	repeat_region	47232..47523	
	Query Match	2.8%; Score 71; DB 9; Length 167531;	
	Best Local Similarity	100.0%; Pred. No. 1,15-28;	
	Matches	71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	2050	TCGAACCTCTGACCTTCAGGTGATCCACCCACCTTGGCTCCCTCCCAAGTGTGGGATTAAAG	2109
Db	108870	TCGAACCTCTCGAAGCTTCAGGTATCCACCCACCTTGGCTCCCAAGTGTGGGATTAAAG	108929
Oy	2110	GTGTGAGGCAC	2120
Db	108930	GTGTGAGGCAC	108940

RESULT	54
LOCUS	AC018636
DEFINITION	Homo sapiens chromosome 7 clone RP11-20F2, complete sequence.
ACCESSION	AC018636
VERSION	AC018636.4 GI:13899420
KEYWORDS	HTS.
ORGANISM	Homo sapiens..
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 167592) Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D. Direct Submission Unpublished 2 (bases 1 to 167592) Bubb,K.U., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M. Direct Submission Submitted (15-DEC-1999) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA 3 (bases 1 to 167592) Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D. Direct Submission Submitted (01-MAY-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA On May 1, 2001 this sequence version replaced gi:9719752.
TITLE	Genome Center
JOURNAL	Center: University of Washington Genome Center Center Code: UWGC Web site: http://www.genome.washington.edu Contact: uwgc@gsu.washington.edu
COMMENT	----- Project Information -----  Center project name: chr7 Center clone name: RP11-20F2 (dj8173)  ----- Summary Statistics ----- Sequencing vector: plasmid; X52328; 100% of reads Chemistry: Dye-primer-amersham; 94% of reads Chemistry: Dye-terminator EP-amersham; 6% of reads Assembly program: Phrap; version 0.990319 Consensus quality: 28568 bases at least Q40 Consensus quality: 29272 bases at least Q30 Insert size: 191971, 14.7% error; agarose-fp Insert size: 169842, sum-of-contigs Quality coverage: 9.6x in Q20 bases; agarose-fp Quality coverage: 10.9x in Q20 bases; sum-of-contigs  ----- Overlapping Sequences: 5': RP11-15G14 (UWGC:dj536) 3': RP11-786T1 (UWGC:dj8736)  ----- Sequence Quality Assessment:





----- Genome Center

Center: Washington University Genome Sequencing Center  
Center code: WUGSCC

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

----- Summary Statistics -----  
Center point name: H NH0140

Center project name: H\_NH0149003

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

### SOURCE INFORMATION

The RPE1-1 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Moon, P. Y., Zhao, B., Frenken, E., Tatenno, M., Cataneese, J. J. and de Jong, P. J. (1996) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.rgsen.com>) or Pictet de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-90J9; the clone sequenced to the right is RP11-122A7. Actual start of this clone is at base position 1 of RP11-14903; actual end is at base position 175345 of RP11-14903.

Data from AC027560 and AC068692 was used to finish this clone AC073065. Polymorphisms have been identified between AC073065 and AC068692.

## FEATURES

```

source
1. .175345
/organism="Homo sapiens"
/db xref="caxon:9606"
/chromosome="2"
/map="2"
/clome="RP11-14903"
/clome_11b="RP11-11"
57. .190
/rpc_family="MIR"
533. .824
/rpc_family="Alu"
905. .1222
/rpc_family="Alu"
1234. .1270
/rpc_family="r-rich"
1242. .1553
/rpc_family="Alu"
1687. .1842
/rpc_family="Alu"
1833. .1893
/rpc_family="AT_rich"
1667. .2288
/rpc_family="L1"
3131. .3159
/rpc_family="L1"

```

```

repeat_region      3160. 3445      /rpt_family="Alu"
repeat_region      3446. 3933      /rpt_family="L1"
repeat_region      3867. 3888      /rpt_family="AT_rich"
repeat_region      4317. 4628      /rpt_family="Alu"
repeat_region      4600. 4645      /rpt_family="Alu"
repeat_region      4677. 4698      /rpt_family="(A)n"
repeat_region      4835. 4859      /rpt_family="AT_rich"
repeat_region      6479. 6675      /rpt_family="(T)n"
repeat_region      6690. 6717      /rpt_family="L2"
repeat_region      6699. 6998      /rpt_family="AT_rich"
repeat_region      7004. 7242      /rpt_family="Alu"
repeat_region      7241. 7282      /rpt_family="L2"
repeat_region      7314. 7414      /rpt_family="(TC)n"
repeat_region      7415. 7629      /rpt_family="L2"
repeat_region      7488. 7491      /rpt_family="Alu"
misc_feature       /note="match to EST A1735669 (NTD:95057193) at19d06_x1"
repeat_region      7609. 7629      /rpt_family="(A)n"
repeat_region      7630. 7665      /rpt_family="L2"
repeat_region      8012. 8143      /rpt_family="Alu"
repeat_region      8434. 8731      /rpt_family="Alu"
repeat_region      8840. 8878      /rpt_family="(CAA)n"
repeat_region      9871. 9919      /rpt_family="(TTTC)n"
repeat_region      9893. 10199     /rpt_family="Alu"
repeat_region      10228. 10529     /rpt_family="Alu"
repeat_region      10508. 10529     /rpt_family="(A)n"
repeat_region      10941. 11293      /rpt_family="MaLR"
repeat_region      11463. 11808      /rpt_family="ERV_L"
repeat_region      11957. 12446      /rpt_family="ERV1"
repeat_region      12571. 12657      /rpt_family="MTR"
repeat_region      12657. 12691      /rpt_family="(T)n"
repeat_region      12665. 12860      /rpt_family="Alu"
repeat_region      13059. 13357      /rpt_family="Alu"
repeat_region      13338. 13357      /rpt_family="(A)n"
repeat_region      13643. 13672      /rpt_family="AT_rich"
repeat_region      13667. 13942      /rpt_family="Alu"
repeat_region      14255. 14603      /rpt_family="L1"
repeat_region      14370. 14393      /rpt_family="AT_rich"
repeat_region      14604. 14901

```

```

repeat_region      /rpt_family="Alu"
                    14902..16152
                    /rpt_family="L1"
repeat_region      16154..16465
                    /rpt_family="Alu"
repeat_region      16470..17053
                    /rpt_family="L1"
repeat_region      17054..17347
                    /rpt_family="Alu"
repeat_region      17348..17572
                    /rpt_family="L1"
repeat_region      17549..17660
                    /rpt_family="L1"
repeat_region      17652..17676
                    /rpt_family="(?)n"
repeat_region      17684..18078
                    /rpt_family="L1"

```

```

Query Match      2.8%; Score 71; DB 9; Length 175345;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2050 TCGAATCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGATTACAG 2109
Db 34099 TCGAATCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGATTACAG 34040
Qy 2110 GTGTGAGCCAC 2120
Db 34039 GTGTGAGCCAC 34029

```

```

RESULT 57
AL391095      178976 bp   DNA   linear   HTG 17-AUG-2000
LOCUS      Homo sapiens chromosome 20 clone RP13-218F6, *** SEQUENCING IN
DEFINITION      AL391095
ACCESSION      AL391095
VERSION      AL391095.5
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 178976)
Direct Submission
Submitted (16-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgen@sanger.ac.uk
requester: clonerequest@sanger.ac.uk
On Aug 18, 2000 this sequence version replaced gi:9798357.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgen@sanger.ac.uk
----- Project Information
Center project name: DB218F6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: MJ3; M7815; 5% of reads
Sequencing vector: plasmid; L08752; 94% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 17658 bases at least Q40
Consensus quality: 17623 bases at least Q20
Insert size: 178876; sum-of-contigs
Insert size: 184221; 3.9% error; agarose-fp
Quality coverage: 13.33x in Q20 bases; sum-of-contigs Quality
Coverage: 12.95x in Q20 bases; agarose-fp

```

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 112029: contig of 112029 bp in length
* 112030 112129: gap of 100 bp
* 112130 178976: contig of 66847 bp in length.
Location/Qualifiers
source
1..178976
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP13-218F6"
/clone_1db="RP13-1"
1..112029
/note="assembly_fragment:00848
clone_end:SP6
vector_side:left"
misc_feature
112130..178976
/note="assembly_fragment:01607"
BASE COUNT 47478 a 41110 c 40786 g 49501 t 101 others
ORIGIN

```

```

Query Match      2.8%; Score 71; DB 2; Length 178976;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2050 TCGAATCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGATTACAG 2109
Db 160413 TCGAATCTCTGACCTGAGTGATCCACCCACTTGGCTCCCAAGTCTGGATTACAG 160472
Qy 2110 GTGTGAGCCAC 2120
Db 160473 GTGTGAGCCAC 160483

```

```

RESULT 58
AC103753      179822 bp   DNA   linear   PRI 25-JUL-2002
LOCUS      Homo sapiens chromosome 15, clone RP11-460N1, complete sequence.
DEFINITION      AC103753
ACCESSION      AC103753
VERSION      AC103753.2
KEYWORDS      HTG.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 179822)
Homo sapiens chromosome 15, clone RP11-460N1
Unpublished
2 (bases 1 to 179822)
REFERENCE
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
TITLE      Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhvalter,B.,
JOURNAL      Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
AUTHORS      Chappel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
TITLE      Cooke,P., D'Arcangelo,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
JOURNAL      Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gerdyta,S.,
AUTHORS      Ginde,S., Gord,S., Goyette,M., Graham,W., Grand-Pierre,N.,
TITLE      Hagos,C., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
JOURNAL      Jones,C., Kamet,A., Karatas,A., Kells,C., Laroque,K.,
AUTHORS      Lamazeres,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
TITLE      Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
JOURNAL      McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrum,J.,
AUTHORS      Menais,L., Mihova,T., Mlenga,V., Murphy,T., Nayler,J., Nguyen,C.,
TITLE      Norou,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
JOURNAL      Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V.,
AUTHORS      Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
TITLE      Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
JOURNAL      Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
AUTHORS      Struss,N., Subramanian,A., Talamas,J., Teafaye,S., Theodore,J.,
TITLE      Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,

```



TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 179822)

Barna, N., Bastein, V., Bloom, T., Boguslavsky, L., Bokksgaard, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Mayor, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Submitted (25-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 25, 2002 this sequence version replaced g1:17149638.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/M/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRS  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L21706  
Center clone name: 460\_N\_1  
-----

## FEATURES

source  
Location/Qualifiers  
1..179822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="RP11-460N1"  
/clone\_lib="RPC1-11 Human Male BAC"  
699..720  
repeat\_region  
/rpt\_family="AT\_rich"  
complement(1244..1553)  
repeat\_region  
/rpt\_family="AluSx"  
2151..2447  
repeat\_region  
/rpt\_family="AluSc"  
2714..2734  
repeat\_region  
/rpt\_family="AT\_rich"  
5083..5210  
repeat\_region  
/rpt\_family="MIR"  
complement(5850..6138)  
repeat\_region  
/rpt\_family="AluSg"  
6984..7287  
repeat\_region  
/rpt\_family="AluY"  
7350..8393  
repeat\_region  
/rpt\_family="L1MB8"  
8421..8732  
repeat\_region  
/rpt\_family="AluSg"  
8733..8867  
repeat\_region  
/rpt\_family="L1MB3"  
8869..9007  
repeat\_region  
/rpt\_family="FAM"  
9173..9193  
repeat\_region  
/rpt\_family="AT\_rich"  
10104..10133  
repeat\_region

repeat\_region  
/rpt\_family="AT\_rich"  
10964..11212  
/rpt\_family="L3"  
repeat\_region  
/rpt\_family="MIR3"  
11287..11413  
complement(11419..11639)  
repeat\_region  
/rpt\_family="MIR"  
11820..12135  
/rpt\_family="AluY"  
12161..12189  
repeat\_region  
/rpt\_family="AT\_rich"  
12395..12680  
/rpt\_family="AluSx"  
complement(12795..13089)  
repeat\_region  
/rpt\_family="AluUb"  
complement(13176..13483)  
/rpt\_family="AluY"  
complement(13697..13833)  
/rpt\_family="L2"  
13834..14139  
/rpt\_family="AluUb"  
complement(14140..14705)  
repeat\_region  
/rpt\_family="L2"  
14854..15152  
/rpt\_family="AluSx"  
15158..15664  
/rpt\_family="L1MB2"  
15665..15950  
/rpt\_family="AluSx"  
15951..16295  
/rpt\_family="L1MB2"  
16437..16533  
/rpt\_family="L2"  
16686..16994  
/rpt\_family="AluSx"  
17963..18291  
/rpt\_family="AluSg"  
19137..19169  
/rpt\_family="CCG)n"  
19593..19666  
/rpt\_family="C-rich"  
complement(20331..20602)  
/rpt\_family="AluSg"  
complement(20648..20751)  
/rpt\_family="MER104"  
complement(20787..20945)  
repeat\_region  
/rpt\_family="MIR3"  
21671..21723  
/rpt\_family="L2"  
21797..21954  
/rpt\_family="MIR"  
22056..22291  
/rpt\_family="MIR"  
complement(22682..22804)  
/rpt\_family="MIR3"  
22846..22971  
/rpt\_family="MIR3"  
23193..23487  
/rpt\_family="AluUb"  
23488..23538  
/rpt\_family="CA)n"  
complement(23874..24181)  
/rpt\_family="AluSx"  
complement(24908..25019)  
/rpt\_family="MIR"  
complement(25942..26076)  
/rpt\_family="MIR"  
complement(27189..27499)  
/rpt\_family="AluSx"  
27590..27690  
/rpt\_family="L1MC4a"  
complement(27776..28060)  
/rpt\_family="AluO"

repeat\_region complement(28074, .28170)  
/rpt\_family="AluYb"  
repeat\_region 28172, .28379  
/rpt\_family="L1MC4a"  
repeat\_region complement(28764, .28873)  
/rpt\_family="MIR"  
repeat\_region complement(28874, .29160)  
/rpt\_family="AluYb"  
repeat\_region complement(29161, .29314)  
/rpt\_family="MIR"  
repeat\_region complement(29495, .29716)  
/rpt\_family="MIR"  
repeat\_region 30060, .30374  
/rpt\_family="AluYb8"  
repeat\_region complement(30492, .30786)  
/rpt\_family="AluYb"  
repeat\_region complement(31594, .31917)  
/rpt\_family="W57B"  
repeat\_region complement(31935, .32227)  
/rpt\_family="L1ME"

Query Match 2.8%; Score 71; DB 9; Length 179822;  
Best Local Similarity 100.0%; Pred. No. 1,1e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAAGCTCTGAGCTGATGATCCACCACTTGGCTCCCAAGTCTGGGATTACAG 2109  
DB 129966 TCGAAGCTCTGAGCTGATGATCCACCACTTGGCTCCCAAGTCTGGGATTACAG 130025  
QY 2110 GTGTGAGCCAC 2120  
DB 130026 GTGTGAGCCAC 130036

RESULT 59  
AC024036/c 181485 bp DNA linear HTG 20-APR-2000  
LOCUS Homo sapiens chromosome X clone RP11-341b6, \*\*\* SEQUENCING IN  
DEFINITION  
AC024036  
AC024036 AC024036.2 GI:7622514  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 181485)  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 181485)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington  
REFERENCE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Apr 20, 2000 this sequence version replaced gi:7008943.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H NE0341P06  
----- Summary Statistics -----  
Sequencing vector: pIasmid; 100k  
Chemistry: Dye-terminator ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 156467 bases at least Q40  
Consensus quality: 165014 bases at least Q30  
Consensus quality: 168940 bases at least Q20

Insert size: 195000; agarose-fp  
Insert size: 178185; sum-of-contigs  
Quality coverage: 3.16 in Q20 bases; agarose-fp  
Quality coverage: 3.58 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 34 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1138: contig of 1138 bp in length  
\* 1139 1238: gap of unknown length  
\* 1239 2662: contig of 1424 bp in length  
\* 2663 2762: gap of unknown length  
\* 2763 3908: contig of 1146 bp in length  
\* 3909 4008: gap of unknown length  
\* 4009 5264: contig of 1256 bp in length  
\* 5265 5364: gap of unknown length  
\* 5365 7138: contig of 1774 bp in length  
\* 7139 7238: gap of unknown length  
\* 7239 9299: contig of 2061 bp in length  
\* 9300 9399: gap of unknown length  
\* 9400 11320: contig of 1921 bp in length  
\* 11321 11420: gap of unknown length  
\* 11421 12731: contig of 1311 bp in length  
\* 12732 12831: gap of unknown length  
\* 12832 15115: contig of 2284 bp in length  
\* 15116 15215: gap of unknown length  
\* 15216 16974: contig of 1759 bp in length  
\* 16975 17074: gap of unknown length  
\* 17075 19448: contig of 2374 bp in length  
\* 19449 19548: gap of unknown length  
\* 19549 22180: contig of 2642 bp in length  
\* 22191 22290: gap of unknown length  
\* 22291 25193: contig of 2903 bp in length  
\* 25194 25293: gap of unknown length  
\* 25294 28512: contig of 3219 bp in length  
\* 28513 28612: gap of unknown length  
\* 28613 31187: contig of 2575 bp in length  
\* 31188 31287: gap of unknown length  
\* 31288 35183: contig of 3896 bp in length  
\* 35184 35283: gap of unknown length  
\* 35284 38128: contig of 2845 bp in length  
\* 38129 38228: gap of unknown length  
\* 38229 42175: contig of 3947 bp in length  
\* 42176 42275: gap of unknown length  
\* 42276 46188: contig of 3913 bp in length  
\* 46189 46288: gap of unknown length  
\* 46289 50486: contig of 4198 bp in length  
\* 50487 50586: gap of unknown length  
\* 50587 55135: contig of 4549 bp in length  
\* 55136 55235: gap of unknown length  
\* 55236 59426: contig of 4191 bp in length  
\* 59427 59527: gap of unknown length  
\* 59528 66329: contig of 6803 bp in length  
\* 66330 66429: gap of unknown length  
\* 66430 71854: contig of 5425 bp in length  
\* 71855 71954: gap of unknown length  
\* 71955 76177: contig of 4222 bp in length  
\* 76178 76276: gap of unknown length  
\* 76277 83383: contig of 7107 bp in length  
\* 83384 83484: gap of unknown length  
\* 83485 90053: contig of 6570 bp in length  
\* 90054 90153: gap of unknown length  
\* 90154 96595: contig of 6442 bp in length  
\* 96596 96695: gap of unknown length  
\* 96696 104875: contig of 8180 bp in length  
\* 104876 104976: gap of unknown length  
\* 104977 115567: contig of 10552 bp in length  
\* 115568 115667: gap of unknown length

```

* 115668 124344: contig of 8677 bp in length
* 124395 124444: gap of unknown length
* 124445 133574: contig of 9130 bp in length
* 133575 133674: gap of unknown length
* 136675 148741: contig of 15067 bp in length
* 148742 148841: gap of unknown length
* 148842 181485: contig of 32644 bp in length.

```

## FEATURES

```

source
  1. 181485
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="X"
    /clone="RP11-341P6"

```

```

BASE COUNT      48948 a 39001 c 39658 g 50566 t 3312 others
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 1,1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2050 TCGAAGCTCTGACCTGAGTGATCCACCCACCTTGCGCTCCCAAGTGCTGGATTACAG 2109
Db 143991 TCGAAGCTCTGACCTGAGTGATCCACCCACCTTGCGCTCCCAAGTGCTGGATTACAG 143932

```

```

QY 2110 GTGTAGGCCAC 2120
Db 143931 GTGTAGGCCAC 143921

```

```

RESULT 60
CNS01DW2      181532 bp  DNA  linear  PRI 20-AUG-2001
LOCUS         Human chromosome 14 DNA sequence BAC R-134415 of library RPEC1-11
DEFINITION    from chromosome 14 of Homo sapiens (human), complete sequence.
ACCESSION     AL136298.5 GI:15282094
VERSION       AL136298.5
KEYWORDS      HTG; HTGS_ACTIVEFIN.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS       Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
               Brotier,P., Catolico,L., Barde,V., Pelleclet,E., Artiguenave,F.,
               Levy,M., Eckenberg,R., Bruls,T., deBardinis,V., Cruaud,C.,
               Gyapay,G., Saurin,W. and Weissenbach,J.
               Sequencing of the human chromosome 14
               Unpublished
               2 (bases 1 to 181532)
               Genoscope.

```

```

TITLE         Direct Submission
JOURNAL        Submitted (03-AUG-2001) Genoscope - Centre National de Sequencage :
REFERENCE      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
AUTHORS        - Web : www.genoscope.cns.fr)
               On Aug 23, 2001 this sequence version replaced gi:13016586.
JOURNAL        ----- Genome Center
TITLE         Center: Genoscope / Centre National de Sequencage
AUTHORS        Center code: GS
               Web site: http://www.genoscope.cns.fr/
               Contact: Seqref@genoscope.cns.fr

```

```

COMMENT
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-320M16 (AC-AL132988)
Downstream BAC (overlapping the SP6 end) : R-431H16 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.89x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 - 9 :

```

## FEATURES

```

source
  10 - 19 :
  20 - 29 : 1
  30 - 39 : 26
  40 - 49 : 337
  50 - 59 : 4864
  60 - 69 : 13321
  70 - 79 : 16232
  80 - 89 : 25050
  90 - 99 : 43828
  78783
Percentage of bases with a quality value >= 40 : 99 %.
Location/Qualifiers
  1. 181532
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="14"
    /clone="R-134E15"
    /clone_id="RPEC1-11"
    /note="matching EMBL:Z39253
    Rdb:RH44252
    Rdb:RH11250
    dbSTS:STS25945
    Identified using the e-PCR software (G. Schuler)"

```

```

BASE COUNT      57126 a 37894 c 35296 g 51216 t
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 1,1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2050 TCGAAGCTCTGACCTGAGTGATCCACCCACCTTGCGCTCCCAAGTGCTGGATTACAG 2109
Db 103688 TCGAAGCTCTGACCTGAGTGATCCACCCACCTTGCGCTCCCAAGTGCTGGATTACAG 103747

```

```

QY 2110 GTGTAGGCCAC 2120
Db 103748 GTGTAGGCCAC 103758

```

```

RESULT 61
AC010204/C      183113 bp  DNA  linear  HTG 27-JUN-2001
LOCUS         Homo sapiens chromosome 12 clone RP11-66N16, WORKING DRAFT
DEFINITION    SEQUENCE. 3 unordered pieces.
ACCESSION     AC010204
VERSION       AC010204.19 GI:14547731
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

## REFERENCE

```

AUTHORS       Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
               Alshrocks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbata,J.,
               Benton,V., Blum,K., Blankenburg,K., Bonnin,D., Bouck,O.,
               Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P., Buhay,C.,
               Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
               Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
               Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
               Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
               Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
               Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
               Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
               Elhai,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
               Foster,P., Franz,J., Gabist,A., Gao,J., Garcia,A., Garner,T.,
               Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
               Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
               Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
               Hollins,B., Homs,J., Howard,S., Huber,J., Hulik,S., Hume,J.,
               Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
               Kovar,C., Kravovic,J., Kurehi,A., Landry,N., Deal,B., Lewis,J.C.,
               Kovar,C., Kravovic,J., Kurehi,A., Landry,N., Deal,B., Lewis,J.C.,

```

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
Unpublished	2 (bases 1 to 183113)	Worley,K.C.	Submitted Submission	(15-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	On Jun 25, 2001 this sequence version replaced gi:14150373.

Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----

NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))

NOTE: This is a working draft sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	88408: contig of 88408 bp in length
88409	88508: gap of unknown length
88509	145820: contig of 57312 bp in length
145821	145820: gap of unknown length
145921	183133: contig of 37193 bp in length.

```

FEATURES
  source      Location/Qualifiers
             1..18313
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="12"
             /clone="Rp11-65N16"
BASE COUNT   56572 a 37358 c 36405 g 52578 t 200 others
ORIGIN

```

```

Query Match      2.8%; Score 71; DB 2; Length 183113;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2050 TCGAAGCTCTGAGACCTCAGGATCTCAACCCACCTTGCGCTCCCAAAAGTGTGGGATTACAG 2109
      |||||
Db 93561 TCGAAGCTCTGAGACCTCAGGATCTCAACCCACCTTGCGCTCCCAAAAGTGTGGGATTACAG 93592

Qy 2110 GTGTGAGCCAC 2120
      |||||
Db 93591 GTGTGAGCCAC 93581

```

RESULT	62
AC002042	
LOCUS	185994 bp linear HTG-23-NOV-1999
DEFINITION	Homo sapiens clone A-18OG2, ** SEQUENCING IN PROGRESS **, S
ACCSSION	unordered pieces.
VERSION	AC002042
KEYWORDS	AC002042.1 GI:3075381 HTG; HTGS_PHASE1.  Homo sapiens. Homo sapiens. Homo sapiens.
SOURCE	
ORGANISM	

REFERENCE	TITLE	JOURNAL
1 (Pages 1 to 185994)		
Authors		
Letts,B.U., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R., Fulman,U., Mason,T., Crosby,M.L., Barnstead,M., Croft,L., Deslattes,Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S., Elchler,E.E., Harris,P.C., Ventner,C.C. and Adams,M.D.	Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q	Genomics 60 (3), 235-306 (1997)

source	1. 185994				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="A-18062"				
BASE COUNT	51275	a	43334	c	45047
ORIGIN			g	48137	t
					201 others
Query Match	2.8%	Score	71	DB 2	Length 185994



```

repeat_region /rpt_family="L2"
                complement(29775, .29967)
                /rpt_family="MIR3"
repeat_region 32167, .32243
                /rpt_family="L2"
                complement(36701, .37208)
repeat_region /rpt_family="LTR9"
                complement(37274, .37592)
repeat_region /rpt_family="L2"
                complement(37649, .37941)
repeat_region /rpt_family="L2"
                complement(38084, .38160)
repeat_region /rpt_family="MIR"
                complement(38161, .38559)
repeat_region /rpt_family="Tigser2a"
                complement(38560, .38589)
repeat_region /rpt_family="MIR"
                complement(39270, .39382)
repeat_region /rpt_family="MIR"
                complement(40652, .40773)
repeat_region /rpt_family="MERSB"
                complement(41260, .41923)
repeat_region /rpt_family="L2"
                complement(42777, .42809)
repeat_region /rpt_family="(TG)n"
                complement(44555, .44610)
repeat_region /rpt_family="Charlie8"
                complement(44785, .44914)
repeat_region /rpt_family="Charlie8"
                complement(45202, .45462)
repeat_region /rpt_family="AlusG1"
                complement(45541, .45578)
repeat_region /rpt_family="AT_rich"
                complement(45890, .45890)
repeat_region /rpt_family="LIMB4"
                complement(46380, .46650)
repeat_region /rpt_family="LIME"
                complement(46687, .46775)
repeat_region /rpt_family="MERSB"
                complement(46849, .46967)
repeat_region /rpt_family="LIM4"
                complement(46993, .47087)
repeat_region /rpt_family="GA-rich"
                complement(47110, .47923)
repeat_region /rpt_family="LIM4"
                complement(48331, .48339)
repeat_region /rpt_family="AT_rich"
                complement(48340, .48631)
repeat_region /rpt_family="AlusG"
                complement(48127, .48281)
repeat_region /rpt_family="LIMB4"
                complement(49408, .49458)
repeat_region /rpt_family="L2"
                complement(49475, .49705)
repeat_region /rpt_family="MIR"
                complement(49723, .49935)
repeat_region /rpt_family="MIR"
                complement(49981, .50452)
repeat_region /rpt_family="LIM4"
                complement(50521, .50633)
repeat_region /rpt_family="L2"
                complement(51723, .52024)
repeat_region /rpt_family="AluY"
                complement(52263, .52604)
repeat_region /rpt_family="MER7"
                complement(53510, .53709)
repeat_region /rpt_family="L2"
                complement(53994, .54109)

```

Query Match 2.8%; Score 71; DB 9; Length 187340;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2050 TCGAACCCTCCTCAGTGTATCCACCCACCTTGCCCTCCCAAGCTCGATTACAG 2109  
 Db 165831 TCGAACCCTCCTCAGTGTATCCACCCACCTTGCCCTCCCAAGCTCGATTACAG 165890  
 Oy 2110 GTGTGAGCCAC 2120  
 Db 165891 GTGTGAGCCAC 165901

RESULT 64  
 AC074234  
 LOCUS  
 DEFINITION Homo sapiens chromosome 1 clone RP1-251N10, WORKING DRAFT  
 SEQUENCE, 46 unordered pieces.  
 AC074234  
 AC074234.5 GI:9958207  
 VERSION  
 KEYWORDS HTG, HTGS\_PHASE1, HTGS\_DRAFT.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL 2 (bases 1 to 187782)  
 REFERENCE  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JUL-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT On Sep 1, 2000 this sequence version replaced gi:9937807.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WGCSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H NH0251N10  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 100%  
 Sequencing vector: plasmid; 0%  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.950319  
 Consensus quality: 159363 bases at least Q40  
 Consensus quality: 166811 bases at least Q20  
 Consensus quality: 170548 bases at least Q20  
 Insert size: 183000; agarose-fp  
 Insert size: 181643; sum-of-contigs  
 Quality coverage: 3.33 in Q20 bases; agarose-fp  
 Quality coverage: 3.50 in Q20 bases; sum-of-contigs  
 ----- NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 46 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1610 1709: contig of 1609 bp in length  
 \* 1710 3287: contig of 1578 bp in length  
 \* 3288 3387: gap of unknown length  
 \* 3388 5026: contig of 1639 bp in length  
 \* 5027 5127: gap of unknown length  
 \* 5127 6828: contig of 1702 bp in length  
 \* 6828 6928: gap of unknown length  
 \* 6928 8638: contig of 1710 bp in length  
 \* 8638 8738: gap of unknown length  
 \* 8738 9929: contig of 1191 bp in length  
 \* 9930 10029: gap of unknown length

\* 10030 11667: contig of 1638 bp in length  
\* 11668 11767: gap of unknown length  
\* 11768 12894: contig of 1227 bp in length  
\* 12995 13094: gap of unknown length  
\* 13095 15391: contig of 2297 bp in length  
\* 15392 15491: gap of unknown length  
\* 15492 17549: contig of 2058 bp in length  
\* 17550 17649: gap of unknown length  
\* 19100 19100: contig of 1451 bp in length  
\* 19101 19200: gap of unknown length  
\* 19201 20950: contig of 1750 bp in length  
\* 20951 21050: gap of unknown length  
\* 21051 22597: contig of 1547 bp in length  
\* 22598 22697: gap of unknown length  
\* 22698 24935: contig of 2238 bp in length  
\* 24936 25035: gap of unknown length  
\* 25036 27148: contig of 2113 bp in length  
\* 27149 27248: gap of unknown length  
\* 27249 29295: contig of 2047 bp in length  
\* 29296 29395: gap of unknown length  
\* 29396 32904: contig of 3509 bp in length  
\* 32905 33004: gap of unknown length  
\* 33005 34986: contig of 1982 bp in length  
\* 34987 35086: gap of unknown length  
\* 35087 38007: contig of 2921 bp in length  
\* 38008 38107: gap of unknown length  
\* 38108 40741: contig of 2634 bp in length  
\* 40742 40841: gap of unknown length  
\* 40842 44444: contig of 3603 bp in length  
\* 44445 44544: gap of unknown length  
\* 44545 48181: contig of 3637 bp in length  
\* 48182 48281: gap of unknown length  
\* 48282 51543: contig of 3262 bp in length  
\* 51544 51643: gap of unknown length  
\* 51644 55269: contig of 3626 bp in length  
\* 55270 55369: gap of unknown length  
\* 55370 58678: contig of 3309 bp in length  
\* 58679 58778: gap of unknown length  
\* 58779 62391: contig of 3613 bp in length  
\* 62392 62491: gap of unknown length  
\* 62492 65538: contig of 3047 bp in length  
\* 65539 65638: gap of unknown length  
\* 65639 69308: contig of 3670 bp in length  
\* 69309 69408: gap of unknown length  
\* 69409 73476: contig of 4068 bp in length  
\* 73477 73576: gap of unknown length  
\* 73577 78759: contig of 5183 bp in length  
\* 78760 78859: gap of unknown length  
\* 78860 83059: contig of 4200 bp in length  
\* 83060 83159: gap of unknown length  
\* 83160 87616: contig of 4457 bp in length  
\* 87617 87716: gap of unknown length  
\* 87717 92361: contig of 4645 bp in length  
\* 92362 92461: gap of unknown length  
\* 92462 97848: contig of 5387 bp in length  
\* 97849 97948: gap of unknown length  
\* 97949 103924: contig of 5976 bp in length  
\* 103925 104024: gap of unknown length  
\* 104025 110342: contig of 6318 bp in length  
\* 110343 110442: gap of unknown length  
\* 110443 116107: contig of 5665 bp in length  
\* 116108 116207: gap of unknown length  
\* 116208 122197: contig of 5990 bp in length  
\* 122198 122297: gap of unknown length  
\* 122299 127047: contig of 4750 bp in length  
\* 127048 127147: gap of unknown length  
\* 127148 133610: contig of 6463 bp in length  
\* 133611 133710: gap of unknown length  
\* 133711 141207: contig of 7497 bp in length  
\* 141208 141307: gap of unknown length  
\* 141308 148738: contig of 7411 bp in length  
\* 148739 148838: gap of unknown length  
\* 148839 156469: contig of 7631 bp in length

FEATURES  
Source  
1. 187782  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-251N10"  
1. 1609  
/note="assembly\_name:Contig17"  
1710. 3287  
/note="assembly\_name:Contig18"  
3388. 5026  
/note="assembly\_name:Contig19"  
5127. 6828  
/note="assembly\_name:Contig20"  
6929. 8638  
/note="assembly\_name:Contig22"  
8739. 9929  
/note="assembly\_name:Contig24"  
10030. 11667  
/note="assembly\_name:Contig25"  
11768. 12994  
/note="assembly\_name:Contig26"  
13095. 15391  
/note="assembly\_name:Contig27"  
15492. 17549  
/note="assembly\_name:Contig28"  
17650. 19100  
/note="assembly\_name:Contig29"  
19201. 20950  
/note="assembly\_name:Contig30"  
21051. 22597  
/note="assembly\_name:Contig31"  
22698. 24935  
/note="assembly\_name:Contig32"  
25036. 27148  
/note="assembly\_name:Contig33"  
27249. 29295  
/note="assembly\_name:Contig34"  
29396. 32904  
/note="assembly\_name:Contig35"  
33005. 34986  
/note="assembly\_name:Contig36"  
35087. 38007  
/note="assembly\_name:Contig37"  
38108. 40741  
/note="assembly\_name:Contig38"  
40842. 44444  
/note="assembly\_name:Contig39"  
44545. 48181  
/note="assembly\_name:Contig40"  
48282. 51543  
/note="assembly\_name:Contig41"  
51644. 55269  
/note="assembly\_name:Contig42"  
55370. 58678  
/note="assembly\_name:Contig43"  
58779. 62391  
/note="assembly\_name:Contig44"  
62492. 65538  
Query Match 2.8%; Score 71; DB 2; Length 187782;  
Best Local Similarity 100.0%; Pred. No. 1.1e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2050 TCGACTCCGACCGCAGGATCCACCCAGCTTGGCCTCCAAAGTGTGGATTACAG 2109  
Db 28365 TCGACTCCGACCTCAGGTATCCACCCAGCTTGGCCTCCAAAGTGTGGATTACAG 28424

OY 2110 GTGTGAGCCAC 2120  
 DB 28425 GTGTGAGCCAC 28435

## RESULT 65

AC112721 188829 bp DNA linear PRI 29-MAY-2002  
 LOCUS Homo sapiens BAC clone RP11-704F14 from 2, complete sequence.  
 AC112721 AC025878  
 VERSION AC112721.3 GI:20800400  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

## REFERENCE

1 (bases 1 to 188829)  
 Suleston,J.E. and Waterston,R.  
 Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)

## MEDLINE

99063792

9847074

## REFERENCE

2 (bases 1 to 188829)  
 Nguyen,C., Haakensen,W., Dignan,G. and Pearson,C.  
 The sequence of Homo sapiens BAC clone RP11-704F14  
 Unpublished (2001)  
 3 (bases 1 to 188829)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (23-FEB-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## REFERENCE

4 (bases 1 to 188829)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (20-MAR-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## REFERENCE

5 (bases 1 to 188829)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (15-MAY-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## REFERENCE

6 (bases 1 to 188829)  
 Waterston,R.  
 Direct Submission  
 Submitted (29-MAY-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On May 15, 2002 this sequence version replaced gi:19551224.

## COMMENT

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0704F14  
 Drafting Center: WIBR  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this  
 sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RP11-704F14 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,  
 Tatematsu,M., Catecaes,J.J. and de Jong,P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
 and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-370J10; the clone sequenced  
 to the right is RP11-810D14, 2000 bp overlap. Actual start of this  
 clone is at base position 1 of RP11-704F14.

An unresolved homopolymeric run exists between bases 2661 to 2722.  
 Polymorphisms have been identified between RP11-757F10 and  
 AC112721. Data from RP11-757F10 and AC112715 was used to finish  
 this clone, AC112721.

## FEATURES

## source

The sequence of AC025878 has been incorporated into AC112721.

## Location/Qualifiers

1..188829  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-704F14"  
 /clone\_11b="RP11-11"  
 410..717  
 /rpt\_family="Alu"  
 1534..1660  
 /rpt\_family="MIR"  
 1768..1837  
 /rpt\_family="L2"  
 2100..2298  
 /rpt\_family="MIR"  
 2363..2722  
 /rpt\_family="Alu"  
 4948..5101  
 /rpt\_family="CRL"  
 5757..5835  
 /rpt\_family="CRL"  
 6273..6411  
 /rpt\_family="Alu"  
 7273..7436  
 /rpt\_family="MIR"  
 11942..12426  
 /rpt\_family="L1"  
 13038..13157  
 /rpt\_family="ERV1"  
 13158..13462  
 /rpt\_family="Alu"  
 13463..13961  
 /rpt\_family="ERV1"  
 15499..15525  
 /rpt\_family="AT\_rich"  
 17881..18088  
 /rpt\_family="L2"  
 23517..23727  
 /rpt\_family="MIR"  
 23871..24091  
 /rpt\_family="(TG)n"  
 24096..24470  
 /rpt\_family="MaLR"  
 24488..24542



```

repeat_region      /rpt_family="CATR" n"
27400..27462
repeat_region      /rpt_family="MIR"
28647..28780
repeat_region      /rpt_family="MIR"
28974..29307
repeat_region      /rpt_family="Alu"
29347..29377
repeat_region      /rpt_family="AT_rich"
29387..29682
repeat_region      /rpt_family="Alu"
29905..30445
repeat_region      /rpt_family="CRY1"
30462..30753
repeat_region      /rpt_family="Alu"
30776..30800
repeat_region      /rpt_family="AT_rich"
31605..32322
repeat_region      /rpt_family="L2"
32519..32577
repeat_region      /rpt_family="MER103"
34229..34657
repeat_region      /rpt_family="MALR"
34658..34988
repeat_region      /rpt_family="Alu"
34985..35051
repeat_region      /rpt_family="MALR"
39145..39166
repeat_region      /rpt_family="AT_rich"
41247..41546
repeat_region      /rpt_family="Alu"
41769..41854
repeat_region      /rpt_family="CR1"
41861..41917
repeat_region      /rpt_family="CR1"
41930..42101
repeat_region      /rpt_family="Alu"
42102..42124
repeat_region      /rpt_family="TTTTG" n"
42125..42244
repeat_region      /rpt_family="Alu"
42671..42729
repeat_region      /rpt_family="(TA) n"
43160..43230
repeat_region      /rpt_family="MIR"
43558..43945
repeat_region      /rpt_family="L2"
44048..44458
repeat_region      /rpt_family="L2"
44604..44722
repeat_region      /rpt_family="MIR"
50651..50957
repeat_region      /rpt_family="Alu"
56138..56203
repeat_region      /rpt_family="(TAGA) n"
56249..56331
repeat_region      /rpt_family="GA-rich"
56341..56518
repeat_region      /rpt_family="(GGA) n"
56553..56657
repeat_region      /rpt_family="GA-rich"
57763..57871

```

Query Match 2.8%; Score 71; DB 9; Length 188829;  
 Best Local Similarity 100.0%; Pred.No. 1.1e-28;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2050 TCGAATCGGACCTGAGTGATTCACCCACCTTGCGCTCCCAAGTCTGGGATTACG 2109  
 DB 163582 TCGAATCGGACCTGAGTGATTCACCCACCTTGCGCTCCCAAGTCTGGGATTACG 163641

QY 2110 GTGTGAGCCAC 2120  
 |||||

Db 163642 GTGTGAGCCAC 163652

```

RESULT 66
AC019118
LOCUS
DEFINITION Homo sapiens BAC clone RP11-513H7 from 2, complete sequence.
AC019118
VERSION
AC019118.7 GI:18464338
KEYWORDS
HTG.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188914)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL
MEDLINE
99063792
PUBMED
9847074
REFERENCE
2 (bases 1 to 188914)
Bielicki, L., Haakenson, W. and Marquis-Homeyer, C.
The sequence of Homo sapiens BAC clone RP11-513H7
Unpublished (2001)
3 (bases 1 to 188914)
Waterson, R.H.
Direct Submission
Submitted (01-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 188914)
Waterson, R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 188914)
Waterson, R.
Direct Submission
Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 1, 2002 this sequence version replaced gi:14280290.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0513H07

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-51 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,

Tateno, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-618J22. Actual start of this clone is at base position 1 of RP11-513H7, actual end is at base position 188914 of RP11-513H7.

There is a region from 45729 to 45825 that is covered by a single M13 subclone. There is a simple sequence repeat from 15184 to 152156. Polymorphisms have been identified between AC073502 and AC019118. Data from AC073502 was used to finish this clone, AC019118.

#### FEATURES

##### source

1. 188914  
Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-513H7"

/clone\_1b="RP11-11"

1185..1482

/rpt\_family="Alu"

1666..2194

/rpt\_family="L1"

2560..2759

/rpt\_family="L2"

3056..3241

/rpt\_family="L2"

3811..4108

/rpt\_family="Alu"

4574..4783

/rpt\_family="MER1\_type"

6533..6538

/note="match to EST B1460036 (NID:g15250692)"

6669..6925

/rpt\_family="MALR"

6969..7147

/rpt\_family="MALR"

8421..8674

/rpt\_family="L1"

8809..9335

/rpt\_family="L1"

/note="match to EST AL523930 (NID:g12787423)"

9376..9526

/rpt\_family="L1"

9527..10469

/rpt\_family="ERVK"

10470..10481

/rpt\_family="L1"

10668..11063

/rpt\_family="MALR"

/note="match to EST AL523930 (NID:g12787423)"

11056..11718

/note="match to EST AL523931 (NID:g12787424)"

11134..11334

/rpt\_family="L1"

11365..11454

/rpt\_family="ERVU"

11606..11654

/rpt\_family="Mariner"

15104..15279

/rpt\_family="(TG)n"

15171..15243

/rpt\_family="ERV1"

15318..15650

/rpt\_family="(TATG)n"

\* repeat\_region

repeat\_region

15658..16018

/rpt\_family="(TG)n"

repeat\_region

16687..16733

/rpt\_family="(A)n"

repeat\_region

16736..17047

/rpt\_family="Alu"

repeat\_region

17232..17312

/rpt\_family="MIR"

repeat\_region

17602..18319

/rpt\_family="L1"

repeat\_region

18416..18635

/rpt\_family="MIR"

repeat\_region

18642..18948

/rpt\_family="L1"

repeat\_region

19097..19223

/rpt\_family="L1"

repeat\_region

19250..19439

/rpt\_family="MIR"

repeat\_region

19816..20208

/rpt\_family="ERVU"

repeat\_region

20242..20598

/rpt\_family="MALR"

repeat\_region

21201..21237

/rpt\_family="(TGAG)n"

repeat\_region

21313..21421

/rpt\_family="L2"

repeat\_region

22746..23426

/note="match to EST BF511221 (NID:g11594519)"

repeat\_region

23189..23565

/rpt\_family="L1"

repeat\_region

23569..23876

/rpt\_family="Alu"

misc\_feature

23575..23591

/note="match to EST A1815922 (NID:g5431468) au43e07.x1"

misc\_feature

23579..23590

/note="match to EST BB858494 (NID:g10373581)"

repeat\_region

23932..23958

/rpt\_family="AT\_rich"

repeat\_region

24272..24528

/rpt\_family="MALR"

repeat\_region

24771..25584

/rpt\_family="L1"

repeat\_region

25592..25772

/rpt\_family="MER2\_type"

repeat\_region

25773..26245

/rpt\_family="L1"

repeat\_region

26246..26366

/rpt\_family="Alu"

repeat\_region

26371..26664

/rpt\_family="Alu"

repeat\_region

26668..27064

/rpt\_family="L1"

repeat\_region

27075..27205

/rpt\_family="L1"

repeat\_region

27588..27879

/rpt\_family="L1"

repeat\_region

27889..28100

/rpt\_family="L1"

repeat\_region

28169..28203

Query Match

Best Local Similarity 100.0%; Pred. No. 1.1e-28;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.8%; Score 71; DB 9; Length 188914;

Db 160446 TCGAATCTGACTGAGTATCCACACCTTGCTCCCAAGTGTGGATTACAG 2109

Db 2110 GTGTGAGCCAC 2120

Db 160506 GTGTGAGCCAC 160516



all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis, MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>.

**SOURCE INFORMATION:**

The RpiC1-1 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frenken, E., Tatenno, M., Catanesi, J. J. and de Jong, P. J. (1993) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.rgsagen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC068343; the clone sequenced to the right is RP11-488C22, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-368E14; actual end is at base position 197709 of RP11-368E14.

Data from AC012460, AC016701, and AC012664 was used to finish this clone, AC073641.

## FEATURES

UNRES	Location/Qualifiers	repeat_region	7625. .7934
source	1. .197709	/rpt_family="Alu"	7937. .7988
	/organism="Homo sapiens"	repeat_region	7937. .7988
	/db_xref="taxon:9606"	/rpt_family="WAlR"	7986. .8208
	/chromosome="2"	repeat_region	7986. .8208
	/map="2"	/rpt_family="WAlR"	8267. .8494
	/clone="RP11-368B14"	repeat_region	8267. .8494
	/clone_id="RPCT-11"	/rpt_family="Alu"	8981. .9035
repeat_region	1. .69	repeat_region	8981. .9035
	/rpt_family="ERV1"	/rpt_family="AT_rich"	9307. .9606
repeat_region	214. .516	repeat_region	9307. .9606
misc_feature	/rpt_family="Alu"	/rpt_family="Alu"	9638. .9664
	721. .921	/rpt_family="TTTC)n"	9742. .9847
misc_feature	/note="match to EST AW753869 (NID:97668801)"	repeat_region	9742. .9847
	774. .904	/rpt_family="Alu"	9903. .9955
misc_feature	/note="match to EST BG958381 (NID:914376552)"	repeat_region	9903. .9955
	796. .919	/rpt_family="CT-rich"	9957. .10702
repeat_region	/note="match to EST BG958432 (NID:914376603)"	repeat_region	9957. .10702
	813. .951	/rpt_family="L1"	10172. .10222
repeat_region	/rpt_family="Alu"	repeat_region	10172. .10222
	925. .954	/rpt_family="(TG)n"	10602. .10622
repeat_region	/rpt_family="AT_rich"	/rpt_family="AT_rich"	10703. .10872
	1308. .1440	/rpt_family="Alu"	10880. .10995
repeat_region	/rpt_family="WIR"	repeat_region	10880. .10995
	1797. .1959	/rpt_family="Alu"	11080. .11205
repeat_region	/rpt_family="WIR"	repeat_region	11080. .11205
	1969. .2257	/rpt_family="MIR"	11417. .11543
repeat_region	/rpt_family="Alu"	repeat_region	11417. .11543
	2253. .2273	/rpt_family="MIR1_type?"	11525. .11702
misc_feature	/note="similar to Homo sapiens EST BE972557 (NID:910585893)"	repeat_region	11525. .11702
	2259. .2563	/rpt_family="Alu"	12025. .12078
repeat_region	/rpt_family="Alu"	repeat_region	12025. .12078
	2261. .2284	/rpt_family="AT_rich"	12462. .12747
repeat_region	/rpt_family="(TTTC)n"	repeat_region	12462. .12747
	2661. .2682	/rpt_family="Alu"	12738. .12757
repeat_region	/rpt_family="AT_rich"	repeat_region	12738. .12757
	3021. .3046	/rpt_family="(CAAA)n"	12823. .12912
repeat_region	/rpt_family="AT_rich"	repeat_region	12823. .12912
	3127. .3194	/rpt_family="MIR"	
repeat_region	/rpt_family="L2"		
	3185. .3609		



```

misc_feature      /note="assembly_name:Contig31"
60153..67197
/note="assembly_name:Contig32"
misc_feature      67298..76866
/note="assembly_name:Contig33"
misc_feature      76967..88157
/note="assembly_name:Contig34"
misc_feature      88258..100018
/note="assembly_name:Contig35
clone_end:17
vector_side:left"
misc_feature      100119..115544
/note="assembly_name:Contig36"
misc_feature      115645..130010
/note="assembly_name:Contig37"
misc_feature      130111..146228
/note="assembly_name:Contig38
clone_end:596
vector_side:right"
misc_feature      146329..164594
/note="assembly_name:Contig39"
misc_feature      164695..198388
/note="assembly_name:Contig40"
BASE COUNT      52607 a 46213 c 45202 g 52163 t 2203 others
ORIGIN
Query Match      2.8%; Score 71; DB 2; Length 198388;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2050 TCGAAGCTCGACCTGATGATCCACCCACCTTGCTCCCAAGTGTGGATTACAG 2109
DB 7605 TCGAAGCTCGACCTGATGATCCACCCACCTTGCTCCCAAGTGTGGATTACAG 7664

OY 2110 GTGTGAGCCAC 2120
DB 7665 GTGTGAGCCAC 7675

RESULT 70
AC022219/c      198398 bp DNA linear HTG 08-JUN-2000
LOCUS      Homo sapiens chromosome 1 clone RP11-784D3, WORKING DRAFT SEQUENCE.
DEFINITION      5 unordered pieces.
ACCESSION      AC022219
VERSION      AC022219.3 GI:8348145
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      1 (bases 1 to 198398)
JOURNAL      Unpublished
REFERENCE      The sequence of Homo sapiens clone
AUTHORS      2 (bases 1 to 198398)
TITLE      Unpublished
JOURNAL      Waterston, R.H.
REFERENCE      Direct Submission
AUTHORS      Submitted (26-JUN-2000) Genome Sequencing Center, Washington
TITLE      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
COMMENT      On Jun 8, 2000 this sequence version replaced gi:7024199.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUSGC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0784D03
----- Summary Statistics -----
Sequencing vector: M13; 80%
Sequencing vector: plasmid; 20%
Chemistry: Dye-primer ET; 80% of reads

```

```

Chemistry: Dye-terminator Big Dye; 20% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195615 bases at least Q40
Consensus quality: 196594 bases at least Q40
Consensus quality: 196985 bases at least Q20
Insert size: 206000; agarose-ff
Insert size: 197998; sum-of-ctdigs
Quality coverage: 6.83 in Q20 bases; agarose-ff
Quality coverage: 7.13 in Q20 bases; sum-of-ctdigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1756: contig of 1756 bp in length
* 1757
* 1856: gap of unknown length
* 1857
* 11573: contig of 9716 bp in length
* 11573
* 11672: gap of unknown length
* 11673
* 40531: contig of 28859 bp in length
* 40532
* 40632: gap of unknown length
* 40632
* 87853: contig of 47222 bp in length
* 87854
* 87854
* 87954: gap of unknown length
* 87954
* 198398: contig of 110445 bp in length.
Location/Qualifiers
1..198398
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-784D3"
BASE COUNT      57352 a 36224 c 39453 g 64967 t 402 others
ORIGIN
Query Match      2.8%; Score 71; DB 2; Length 198398;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2050 TCGAAGCTCGACCTGATGATCCACCCACCTTGCTCCCAAGTGTGGATTACAG 2109
DB 25700 TCGAAGCTCGACCTGATGATCCACCCACCTTGCTCCCAAGTGTGGATTACAG 25641

OY 2110 GTGTGAGCCAC 2120
DB 25640 GTGTGAGCCAC 25630

RESULT 71.
AC016719/c      198748 bp DNA linear HTG 13-AUG-2000
LOCUS      Homo sapiens chromosome 5 clone RP11-323J7, WORKING DRAFT SEQUENCE.
DEFINITION      4 unordered pieces.
ACCESSION      AC016719
VERSION      AC016719.2 GI:9739335
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      1 (bases 1 to 198748)
JOURNAL      Unpublished
REFERENCE      The sequence of Homo sapiens clone
AUTHORS      2 (bases 1 to 198748)
TITLE      Unpublished
JOURNAL      Waterston, R.H.
REFERENCE      Direct Submission
AUTHORS      Submitted (04-DEC-1999) Genome Sequencing Center, Washington
TITLE      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
COMMENT      On Aug 8, 2000 this sequence version replaced gi:6524345.

```

ACCESSION	anti-oncogene region, section 2/5.
AP000434	
KEYWORDS	AP000434.1 GI:5926671
SOURCE	
ORGANISM	Homo sapiens DNA.
REFERENCE	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (sites)
REFERENCE	Nakamura,Y., Isonuma,M., Daigo,Y., Tamari,M. and Ishikawa,S.
AUTHORS	DNA sequence analysis of a 1.2-Mb region on chromosome 3p11.3
TITLE	Published Only in Database (1999)
JOURNAL	2 (bases 1 to 200000)
REFERENCE	Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
AUTHORS	Direct Submission
TITLE	Submitted (21-SEP-1999) Mika Hirakawa, Japan Science and Technology
JOURNAL	Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
REFERENCE	Chiyo-da-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
AUTHORS	URL: http://www-alls.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
TITLE	Fax:81-3-5214-8170)
JOURNAL	This sequence is conducted by Human Genome Sequencing in ALIS
REFERENCE	Research as a JST sequencing team.
AUTHORS	Principal Investigator: Yusuke Nakamura Ph.D
TITLE	Phone:81-3-5449-5372, Fax:81-3-5449-5433,
JOURNAL	Yusuke@jst.ims.u-tokyo.ac.jp
REFERENCE	The sequence is submitted by Human Genome Sequencing in ALIS
AUTHORS	Project of JST
TITLE	Japan Science and Technology Corporation (JST)
JOURNAL	5-3, Yonbancho, Chiyo-da-ku, Tokyo, 102-0081 Japan
REFERENCE	For further information about this sequences, please visit our
AUTHORS	research archive Web site (http://www-alls.tokyo.jst.go.jp/HGS/top.
TITLE	html) or send email to webmaster@www-alls.tokyo.jst.go.jp.
JOURNAL	Location/Qualifiers
FEATURES	
source	1..200000
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/map="3p21.3"
	/note="301-308"
	complement(112666..112810)
	/note="srsG9202:The location is between each flanking site
	of PCR primers."
	/db_xref="GDB:4569933"
BASE COUNT	52190 a 42809 c 46061 g 58756 t 184 others
ORIGIN	
Query Match	2.8%; Score 71; DB 9; Length 200000;
Best Local Similarity	100.0%; Pred. No. 1,1e-28;
Matches	71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	2050 TCGAAGCTCGAAGCTCGAAGTATCCACCCAGCTTGAGCTCCCAAGTGTGGATTTCAG 2109
Dp	139609 TCGAAGCTCGAAGCTCGAAGTATCCACCCAGCTTGAGCTCCCAAGTGTGGATTTCAG 139668
Qy	2110 GTGAGAGCCAC 2120
Dp	139669 GTGAGAGCCAC 139679
RESULT 73	
LOCUS	AC012460/c
DEFINITION	Homo sapiens chromosome 2 clone RP11-536G14, WORKING DRAFT
ACCESSION	AC012460.5 GI:13027576
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS	Homo sapiens.
SOURCE	Homo sapiens.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 200885)
REFERENCE	Waterston,R.H.

TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 200885)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (27-OCT-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Feb 21, 2001 this sequence version replaced gi:8570287.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: MUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Project Information -----  
Center project name: H\_NH0536G14  
----- Summary Statistics -----  
Sequencing vector: MJ3; 73%  
Sequencing vector: plasmid; 27%  
Chemistry: Dye-terminator Big Dye; 27% of reads  
Chemistry: Dye-terminator Big Dye; 27% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 195782 bases at least Q40  
Consensus quality: 197726 bases at least Q30  
Consensus quality: 198568 bases at least Q20  
Insert size: 20200; agarose-fp  
Insert size: 200085; sum-of-contigs  
Quality coverage: 4.91 in Q20 bases; agarose-fp  
Quality coverage: 5.28 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1894: contig of 1894 bp in length  
\* 1895 1894: gap of unknown length  
\* 1896 8748: contig of 6754 bp in length  
\* 8749 8848: gap of unknown length  
\* 8849 14924: contig of 6076 bp in length  
\* 14925 15024: gap of unknown length  
\* 15025 26086: contig of 11067 bp in length  
\* 26087 26186: gap of unknown length  
\* 26187 38993: contig of 12807 bp in length  
\* 38994 39093: gap of unknown length  
\* 39094 51695: contig of 12602 bp in length  
\* 51696 51795: gap of unknown length  
\* 51796 79679: contig of 27884 bp in length  
\* 79680 79779: gap of unknown length  
\* 79780 112476: contig of 32697 bp in length  
\* 112477 112576: gap of unknown length  
\* 112577 200885: contig of 88309 bp in length.  
FEATURES  
source  
1..200885  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-536G14"  
1..1894  
/note="assembly\_name:Contig20"  
misc\_feature 1895..8748  
/note="assembly\_name:Contig21"  
misc\_feature 8849..14924  
/note="assembly\_name:Contig22"  
vector\_end:17  
vector\_side:right  
15025..26086  
/note="assembly\_name:Contig23"  
26187..38993  
misc\_feature  
/note="assembly\_name:Contig24"

clone\_end:56  
vector\_side:right  
39094..51695  
/note="assembly\_name:Contig25"  
misc\_feature 51796..79679  
/note="assembly\_name:Contig26"  
misc\_feature 79780..112476  
/note="assembly\_name:Contig27"  
misc\_feature 112577..200885  
/note="assembly\_name:Contig28"  
BASE COUNT 56415 a 42233 c 42806 g 58627 t 804 others  
ORIGIN  
Query Match 2.8%; Score 71; DB 2; Length 200885;  
Best Local Similarity 100.0%; Pred. No. 1.1e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2050 TCGAATCTCTGACTGATTCACCCACCTTGCTCCCAAGTGTGGATTACAG 2109  
DB 146180 TCGAATCTCTGACTGATTCACCCACCTTGCTCCCAAGTGTGGATTACAG 146121  
OY 2110 GTGTGAGCCAC 2120  
DB 146120 GTGTGAGCCAC 146110  
RESULT 74  
AC005071/c 209382 bp DNA linear PRI 24-MAR-1999  
LOCUS Homo sapiens clone RG161A02, complete sequence.  
DEFINITION AC005071  
AC005071.2 GI:4508112  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 209382)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 209382)  
Waterston,R.H.  
AUTHORS Direct Submission  
TITLE Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 209382)  
REFERENCE Direct Submission  
AUTHORS Waterston,R.H.  
TITLE Submitted (24-MAR-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Mar 24, 1999 this sequence version replaced gi:3212920.  
FEATURES  
source  
1..209382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RG161A02"  
BASE COUNT 53274 a 51565 c 52404 g 52139 t  
ORIGIN  
Query Match 2.8%; Score 71; DB 9; Length 209382;  
Best Local Similarity 100.0%; Pred. No. 1.1e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2050 TCGAATCTCTGACTGATTCACCCACCTTGCTCCCAAGTGTGGATTACAG 2109  
DB 193758 TCGAATCTCTGACTGATTCACCCACCTTGCTCCCAAGTGTGGATTACAG 193699  
OY 2110 GTGTGAGCCAC 2120  
DB 193698 GTGTGAGCCAC 193688



RESULT 75	LOCUS	211945 bp	DNA	linear	PRI 31-JAN-1996
AC006238	AC006238	211945 bp	DNA	linear	PRI 31-JAN-1996
DEFINITION	Homo sapiens chromosome 18, clone hRPK.474_N_24, complete sequence.				
ACCESSION	AC006238				
VERSION	AC006238.1	GI:4204704			
KEYWORDS	HTG.				
ORGANISM	Homo sapiens.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 211945)				
TITLE	Biren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens chromosome 18, clone hRPK.474_N_24				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 211945)				
TITLE	Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckely,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Harford,A., Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,M., Morrow,J., Mychalecky,J., Naylor,J., Nioffe,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.				
JOURNAL	Submitted				
REFERENCE	Submitted (18-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
AUTHORS	3 (bases 1 to 211945)				
TITLE	Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckely,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Harford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,M., Morrow,J., Mychalecky,J., Naylor,J., Nioffe,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (18-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
AUTHORS	3 (bases 1 to 211945)				
TITLE	Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckely,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Harford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,M., Morrow,J., Mychalecky,J., Naylor,J., Nioffe,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (31-JAN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
AUTHORS	On Jan 31, 1999 this sequence version replaced gi:190946.				
TITLE	All repeats were identified using RepeatMasker: smt, A.F.A. & Green, P. (1996-1997)				
JOURNAL	http://ftp.genome.washington.edu/RW/RepeatMasker.html.				
REFERENCE	Location/Qualifiers				
AUTHORS	1. 211945				
TITLE	/organism="Homo sapiens"				
JOURNAL	/db_xref="taxon:9606"				
REFERENCE	/chromosome="18"				
AUTHORS	/map="18"				
TITLE	/clone="hRPK.474_N_24"				
JOURNAL	/clone_lib="RPCT-IT human BAC library"				
REFERENCE	/complement(32..1833)				
AUTHORS	/rptc_family="LIM3c"				
TITLE	2108..2211				
JOURNAL	/rptc_family=" (TG) n"				
REFERENCE	3040..3092				
AUTHORS	/rptc_family=" (CAGT) n"				
TITLE	complement(6746..7851)				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat_region				
AUTHORS	repeat_region				
TITLE	repeat_region				
JOURNAL	repeat_region				
REFERENCE	repeat				

Job time : 17913 secs

repeat_region	399.6 , .40082
/rpt_family="MIR"	4049.5 , .40637
repeat_region	/rpt_family="MEMB5B"
repeat_region	4178.3 , .42073
repeat_region	/rpt_family="Alusg"
repeat_region	4285.6 , .42965
repeat_region	/rpt_family="MIR"
repeat_region	451.20 , .45328
repeat_region	/rpt_family="LMB8A"
repeat_region	4624.6 , .46398
repeat_region	/rpt_family="MEMB5B"
repeat_region	complement(4650.0 , .46819)
repeat_region	/rpt_family="Alusg"
repeat_region	complement(479.1 , .48228)
repeat_region	/rpt_family="MER33"
repeat_region	5024.1 , .50533
repeat_region	/rpt_family="AluDo"
repeat_region	511.8 , .51215
repeat_region	/rpt_family="(TA)n"
repeat_region	516.24 , .52514
repeat_region	/rpt_family="Tigger3b"
repeat_region	525.15 , .52741
repeat_region	/rpt_family="MEMR30"
repeat_region	5274.2 , .52934
repeat_region	/rpt_family="Tigger3b"
repeat_region	5334.2 , .53375
repeat_region	/rpt_family="AT_rich"
repeat_region	5345.0 , .53754
repeat_region	/rpt_family="AlusX"
repeat_region	5495.7 , .55256
repeat_region	/rpt_family="AluDoB"
repeat_region	5588.8 , .55921
repeat_region	/rpt_family="AT_rich"
repeat_region	5616.8 , .56581
repeat_region	/rpt_family="MGTA"
repeat_region	5683.3 , .56920
repeat_region	/rpt_family="LMNA9"
repeat_region	5796.2 , .58106
repeat_region	/rpt_family="MEMR103"
repeat_region	5870.3 , .59203
repeat_region	/rpt_family="LMMS"
repeat_region	complement(5921.2 , .59426)
repeat_region	/rpt_family="MEMR20"
repeat_region	5944.3 , .59650
repeat_region	/rpt_family="LMMS"
repeat_region	5963.1 , .59937
repeat_region	/rpt_family="AlusP"
repeat_region	5993.8 , .60333
repeat_region	/rpt_family="LMMS"
repeat_region	6186.6 , .61866
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(6237.0 , .62523)
repeat_region	/rpt_family="AluDo"
repeat_region	6252.4 , .62557
repeat_region	/rpt_family="(TTTG)n"
repeat_region	complement(6255.8 , .62676)
repeat_region	/rpt_family="AluDo"

```
Query Match      2.8%; Score 71; DB 9; Length 211945;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

[illegible]

```

QY      2110 GTGTGAGCCAC 2120
          |||||
Db 109298 GTGTGAGCCAC 109308

```

Search completed: March 30, 2003, 21:28:58